L4: Entry 9 of 11 File: USPT Mar 12, 2002

DOCUMENT-IDENTIFIER: US 6355459 B1
TITLE: Genes for the biosynthesis of epothilones

Abstract Text (1):
Nucleic acid molecules are isolated from Sorangium cellulosum that encode polypeptides necessary for the biosynthesis of epothilone. Disclosed are methods for the production of epothilone in recombinant hosts transformed with the genes of the invention. In this manner, epothilone can be produced in quantities large enough to enable their purification and use in pharmaceutical formulations such as those for the treatment of cancer.

Brief Summary Text (2):
The present invention relates generally to polyketides and genes for their synthesis. In particular, the present invention relates to the isolation and characterization of novel polyketide synthase and nonribosomal peptide synthetase genes from Sorangium cellulosum that are necessary for the biosynthesis of epothilones A and B.

Brief Summary Text (4):
Polyketides are compounds synthesized from two-carbon building blocks, the .beta.-carbon of which always carries a keto group, thus the name polyketide. These compounds include many important antibiotics, immunosuppressants, cancer chemotherapeutic agents, and other compounds possessing a broad range of biological properties. The tremendous structural diversity derives from the different lengths of the polyketide chain, the different side-chains introduced (either as part of the two-carbon building blocks or after the polyketide backbone is formed), and the stereochemistry of such groups. The keto groups may also be reduced to hydroxyls, enoyls, or removed altogether. Each round of two-carbon addition is carried out by a complex of enzymes called the polyketide synthase (PKS) in a manner similar to fatty acid biosynthesis.

Brief Summary Text (5):
The biosynthetic genes for an increasing number of polyketides have been isolated and sequenced. For example, see U.S. Pat. Nos. 5,639,949, 5,693,774, and 5,716,849, all of which are incorporated herein by reference, which describe genes for the biosynthesis of soraphen. See also, Schupp et al., FEMS Microbiology Letters 159: 201-207 (1998) and WO 98/07868, which describe genes for the biosynthesis of rifamycin, and U.S. Pat. No. 5,876,991, which describes genes for the biosynthesis of tylactone, all of which are incorporated herein by reference. The encoded proteins generally fall into two types: type I and type II. Type I proteins are polyfunctional, with several catalytic domains carrying out different enzymatic steps covalently linked together (e.g. PKS for erythromycin, soraphen, rifamycin, and avermectin (MacNeil et al., in Industrial Microorganisms: Basic and Applied Molecular Genetics, (ed.: Baltz et al.), American Society for Microbiology, Washington D.C. pp. 245-256 (1993)); whereas type II proteins are monofunctional (Hutchinson et al., in Industrial Microorganisms: Basic and Applied Molecular Genetics, (ed.: Baltz et al.), American Society for Microbiology, Washington D.C. pp. 203-216 (1993)).

Brief Summary Text (6):
For the simpler polyketides such as actinorhodin (produced by Streptomyces coelicolor), the several rounds of two-carbon additions are carried out iteratively on PKS enzymes encoded by one set of PKS genes. In contrast, synthesis of the more complicated compounds such as erythromycin and soraphen involves PKS enzymes that are organized into modules, whereby each module carries out one round of two-carbon addition (for review, see Hopwood et al., in Industrial Microorganisms: Basic and Applied Molecular Genetics, (ed.: Baltz et al.), American Society for Microbiology, Washington D.C., pp. 267-275 (1993)).

Brief Summary Text (8):
Epothilones A and B are 16-membered macrocyclic polyketides with an acylcysteine-derived starter unit that are produced by the bacterium Sorangium cellulosum strain So ce90 (Gerth et al., J. Antibiotics 49: 560-563 (1996), incorporated herein by reference). The structure of epothilone A and B wherein R signifies hydrogen (epothilone A) or methyl (epothilone B) is:
##STR1##

Brief Summary Text (9):
The epothilones have a narrow antifungal spectrum and especially show a high cytotoxicity in animal cell cultures (see, Hofle et al., Patent DE 4138042 (1993), incorporated herein by reference). Of significant importance, epothilones mimic the biological effects of taxol, both in vivo and in cultured cells (Bollag et al., Cancer Research 55: 2325-2333 (1995), incorporated herein by reference). Taxol and taxotere, which stabilize cellular microtubules, are cancer chemotherapeutic agents with significant activity against various human solid tumors (Rowinsky et al., J. Natl. Cancer Inst. 83: 1778-1781 (1991)). Competition studies have revealed that epothilones act as competitive inhibitors of taxol binding to microtubules, consistent with the interpretation that they share the same microtubule-binding site and possess a similar microtubule affinity as taxol. However, epothilones enjoy a significant advantage over taxol in that epothilones exhibit a much lower drop in potency compared to taxol against a multiple drug-resistant cell line (Bollag et al. (1995)). Furthermore, epothilones are considerably less efficiently exported from the cells by P-glycoprotein than is taxol (Gerth et al. (1996)). In addition, several epothilone analogs have been synthesized that have a superior cytotoxic activity as compared to epothilone A or epothilone B as demonstrated by their enhanced ability to induce the polymerization and stabilization of microtubules (WO 98/25929, incorporated herein by reference).

Brief Summary Text (10):
Despite the promise shown by the epothilones as anticancer agents, problems pertaining to the production of these compounds presently limit their commercial potential. The compounds are too complex for industrial-scale chemical synthesis and so must be produced by fermentation. Techniques for the genetic manipulation of myxobacteria such as Sorangium cellulosum are described in U.S. Pat. No. 5,686,295, incorporated herein by reference. However, Sorangium cellulosum is notoriously difficult to ferment and production levels of epothilones are therefore low. Recombinant production of epothilones in heterologous hosts that are more amenable to fermentation could solve current production problems. However, the genes that encode the polypeptides responsible for epothilone bio-synthesis have heretofore not been isolated. Furthermore, the strain that produces epothilones, i.e. So ce90, also produces at least one additional polyketide, spirangien, which would be expected to greatly complicate the isolation of the genes particularly responsible for epothilone biosynthesis.

Brief Summary Text (11):
Therefore, in view of the foregoing, one object of the present invention is to isolate the genes that are involved in the synthesis of epothilones, particularly the genes that are involved in the synthesis of epothilones A and B in myxobacteria of the Sorangium/Polyangium group, i.e., Sorangium cellulosum strain So ce90. A further object of the invention is to provide a method for the recombinant production of epothilones for application in anticancer formulations.

Brief Summary Text (13):
In furtherance of the aforementioned and other objects, the present invention unexpectedly overcomes the difficulties set forth above to provide for the first time a nucleic acid molecule comprising a nucleotide sequence that encodes at least one polypeptide involved in the biosynthesis of epothilone. In a preferred embodiment, the nucleotide sequence is isolated from a species belonging to Myxobacteria, most preferably Sorangium cellulosum.

Brief Summary Text (14):
In another preferred embodiment, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes at least one polypeptide involved in the biosynthesis of an epothilone, wherein said polypeptide comprises an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: SEQ ID NO:2, amino acids 11-437 of SEQ ID NO:2, amino acids 543-864 of SEQ ID NO:2, amino acids 974-1273 of SEQ ID NO:2, amino acids 1314-1385 of SEQ ID NO:2, SEQ ID NO:3, amino acids 72-81 of SEQ ID NO:3, amino acids 118-125 of SEQ ID NO:3, amino acids 199-212 of SEQ ID NO:3, amino acids 353-363 of SEQ ID NO:3, amino acids 549-565 of SEQ ID NO:3, amino acids 588-603 of SEQ ID NO:3, amino acids 669-684 of SEQ ID NO:3, amino acids 815-821 of SEQ ID NO:3, amino acids 868-892 of SEQ ID NO:3, amino acids 903-912 of SEQ ID NO:3, amino acids 918-940 of SEQ ID NO:3, amino acids 1268-1274 of SEQ ID NO:3, amino acids 1285-1297 of SEQ ID NO:3, amino acids 973-1256 of SEQ ID NO:3, amino acids 1344-1351 of SEQ ID NO:3, SEQ ID NO:4, amino acids 7-432 of SEQ ID NO:4, amino acids 539-859 of SEQ ID NO:4, amino acids 869-1037 of SEQ ID NO:4, amino acids 1439-1684 of SEQ ID NO:4, amino acids 1722-1792 of SEQ ID NO:4, SEQ ID NO:5, amino acids 39-457 of SEQ ID NO:5, amino acids 563-884 of SEQ ID NO:5, amino acids 1147-1399 of SEQ ID NO:5, amino acids 1434-1506 of SEQ ID NO:5, amino acids 1524-1950 of SEQ ID NO:5, amino acids 2056-2377 of SEQ ID NO:5, amino acids 2645-2895 of SEQ ID NO:5, amino acids 2932-3005 of SEQ ID NO:5, amino acids 3024-3449 of SEQ ID NO:5, amino acids 3555-3876 of SEQ ID NO:5, amino acids 3886-4048 of SEQ ID NO:5, amino acids 4433-4719 of SEQ ID NO:5, amino acids 4729-4974 of SEQ ID NO:5, amino acids 5010-5082 of SEQ ID NO:5, amino acids 5103-5525 of SEQ ID NO:5, amino acids 5631-5951 of SEQ ID NO:5, amino acids 5964-6132 of SEQ ID NO:5, amino acids 6542-6837 of SEQ ID NO:5, amino acids 6857-7101 of SEQ ID NO:5, amino acids 7140-7211 of SEQ ID NO:5, SEQ ID NO:6, amino acids 35-454 of SEQ ID NO:6, amino acids 561-881 of SEQ ID NO:6, amino acids 1143-1393 of SEQ ID NO:6, amino acids 1430-1503 of SEQ ID NO:6, amino acids 1522-1946 of SEQ ID NO: 6, amino acids 2053-2373 of SEQ ID NO:6, amino acids 2383-2551 of SEQ ID NO:6, amino acids 2671-3045 of

SEQ ID NO:6, amino acids 3392-3636 of SEQ ID NO:6, amino acids 3673-3745 of SEQ ID NO:6, SEQ ID NO:7, amino acids 32-450 of SEQ ID NO:7, amino acids 556-877 of SEQ ID NO:7, amino acids 887-1051 of SEQ ID NO:7, amino acids 1478-1790 of SEQ ID NO:7, amino acids 1810-2055 of SEQ ID NO:7, amino acids 2093-2164 of SEQ ID NO:7, amino acids 2165-2439 of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, and SEQ ID NO:22.

Brief Summary Text (15):
In a more preferred embodiment, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes at least one polypeptide involved in the biosynthesis of an epothilone, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: SEQ ID NO:2, amino acids 11-437 of SEQ ID NO:2, amino acids 543-864 of SEQ ID NO:2, amino acids 974-1273 of SEQ ID NO:2, amino acids 1314-1385 of SEQ ID NO:2, SEQ ID NO:3, amino acids 72-81 of SEQ ID NO:3, amino acids 118-125 of SEQ ID NO:3, amino acids 199-212 of SEQ ID NO:3, amino acids 353-363 of SEQ ID NO:3, amino acids 549-565 of SEQ ID NO:3, amino acids 588-603 of SEQ ID NO:3, amino acids 669-684 of SEQ ID NO:3, amino acids 815-821 of SEQ ID NO:3, amino acids 868-892 of SEQ ID NO:3, amino acids 903-912 of SEQ ID NO:3, amino acids 918-940 of SEQ ID NO:3, amino acids 1268-1274 of SEQ ID NO:3, amino acids 1285-1297 of SEQ ID NO:3, amino acids 973-1256 of SEQ ID NO:3, amino acids 1344-1351 of SEQ ID NO:3, SEQ ID NO:4, amino acids 7-432 of SEQ ID NO:4, amino acids 539-859 of SEQ ID NO:4, amino acids 869-1037 of SEQ ID NO:4, amino acids 1439-1684 of SEQ ID NO:4, amino acids 1722-1792 of SEQ ID NO:4, SEQ ID NO:5, amino acids 39-457 of SEQ ID NO:5, amino acids 563-884 of SEQ ID NO:5, amino acids 1147-1399 of SEQ ID NO:5, amino acids 1434-1506 of SEQ ID NO:5, amino acids 1524-1950 of SEQ ID NO:5, amino acids 2056-2377 of SEQ ID NO:5, amino acids 2645-2895 of SEQ ID NO:5, amino acids 2932-3005 of SEQ ID NO:5, amino acids 3024-3449 of SEQ ID NO:5, amino acids 3555-3876 of SEQ ID NO:5, amino acids 3886-4048 of SEQ ID NO:5, amino acids 4433-4719 of SEQ ID NO:5, amino acids 4729-4974 of SEQ ID NO:5, amino acids 5010-5082 of SEQ ID NO:5, amino acids 5103-5525 of SEQ ID NO:5, amino acids 5631-5951 of SEQ ID NO:5, amino acids 5964-6132 of SEQ ID NO:5, amino acids 6542-6837 of SEQ ID NO:5, amino acids 6857-7101 of SEQ ID NO:5, amino acids 7140-7211 of SEQ ID NO:5, SEQ ID NO:6, amino acids 35-454 of SEQ ID NO:6, amino acids 561-881 of SEQ ID NO:6, amino acids 1143-1393 of SEQ ID NO:6, amino acids 1430-1503 of SEQ ID NO:6, amino acids 1522-1946 of SEQ ID NO: 6, amino acids 2053-2373 of SEQ ID NO:6, amino acids 2383-2551 of SEQ ID NO:6, amino acids 2671-3045 of SEQ ID NO:6, amino acids 3392-3636 of SEQ ID NO:6, amino acids 3673-3745 of SEQ ID NO:6, SEQ ID NO:7, amino acids 32-450 of SEQ ID NO:7, amino acids 556-877 of SEQ ID NO:7, amino acids 887-1051 of SEQ ID NO:7, amino acids 1478-1790 of SEQ ID NO:7, amino acids 1810-2055 of SEQ ID NO:7, amino acids 2093-2164 of SEQ ID NO:7, amino acids 2165-2439 of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, and SEQ ID NO:22.

Brief Summary Text (16):
In yet another preferred embodiment, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes at least one polypeptide involved in the biosynthesis of an epothilone, wherein said nucleotide sequence is substantially similar to a nucleotide sequence selected from the group consisting of: the complement of nucleotides 1900-3171 of SEQ ID NO:1, nucleotides 3415-5556 of SEQ ID NO:1, nucleotides 7610-11875 of SEQ ID NO:1, nucleotides 7643-8920 of SEQ ID NO:1, nucleotides 9236-10201 of SEQ ID NO:1, nucleotides 10529-11428 of SEQ ID NO:1, nucleotides 11549-11764 of SEQ ID NO:1, nucleotides 11872-16104 of SEQ ID NO:1, nucleotides 12085-12114 of SEQ ID NO:1, nucleotides 12223-12246 of SEQ ID NO:1, nucleotides 12466-12507 of SEQ ID NO:1, nucleotides 12928-12960 of SEQ ID NO:1, nucleotides 13516-13566 of SEQ ID NO:1, nucleotides 13633-13680 of SEQ ID NO:1, nucleotides 13876-13923 of SEQ ID NO:1, nucleotides 14313-14334 of SEQ ID NO:1, nucleotides 14473-14547 of SEQ ID NO:1, nucleotides 14578-14607 of SEQ ID NO:1, nucleotides 14623-14692 of SEQ ID NO:1, nucleotides 15673-15693 of SEQ ID NO:1, nucleotides 15724-15762 of SEQ ID NO:1, nucleotides 14788-15639 of SEQ ID NO:1, nucleotides 15901-15924 of SEQ ID NO:1, nucleotides 16251-21749 of SEQ ID NO:1, nucleotides 16269-17546 of SEQ ID NO:1, nucleotides 17865-18827 of SEQ ID NO:1, nucleotides 18855-19361 of SEQ ID NO:1, nucleotides 20565-21302 of SEQ ID NO:1, nucleotides 21414-21626 of SEQ ID NO:1, nucleotides 21746-43519 of SEQ ID NO:1, nucleotides 21860-23116 of SEQ ID NO:1, nucleotides 23431-24397 of SEQ ID NO:1, nucleotides 25184-25942 of SEQ ID NO:1, nucleotides 26045-26263 of SEQ ID NO:1, nucleotides 26318-27595 of SEQ ID NO:1, nucleotides 27911-28876 of SEQ ID NO:1, nucleotides 29678-30429 of SEQ ID NO:1, nucleotides 30539-30759 of SEQ ID NO:1, nucleotides 30815-32092 of SEQ ID NO:1, nucleotides 32408-33373 of SEQ ID NO:1, nucleotides 33401-33889 of SEQ ID NO:1, nucleotides 35042-35902 of SEQ ID NO:1, nucleotides 35930-36667 of SEQ ID NO:1, nucleotides 36773-36991 of SEQ ID NO:1, nucleotides 37052-38320 of SEQ ID NO:1, nucleotides 38636-39598 of SEQ ID NO:1, nucleotides 39635-40141 of SEQ ID NO:1, nucleotides 41369-42256 of SEQ ID NO:1, nucleotides 42314-43048 of SEQ ID NO:1, nucleotides 43163-43378 of SEQ ID NO:1, nucleotides 43524-54920 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, nucleotides 51534-52657 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, nucleotides 54540-54758 of SEQ ID NO:1, nucleotides 54935-62254 of SEQ ID NO:1, nucleotides 55028-56284 of SEQ ID NO:1, nucleotides 56600-57565 of SEQ ID NO:1, nucleotides 57593-58087 of SEQ ID NO:1, nucleotides 59366-60304 of SEQ ID NO:1, nucleotides 60362-61099 of SEQ ID NO:1, nucleotides 61211-61426 of SEQ ID NO:1, nucleotides 61427-62254 of SEQ ID NO:1, nucleotides 62369-63628 of SEQ ID NO:1, nucleotides 67334-68251 of SEQ ID NO:1, and nucleotides 1-68750 SEQ ID NO:1.

Brief Summary Text (17):
In an especially preferred embodiment, the present invention provides a nucleic acid molecule comprising a nucleotide sequence that encodes at least one polypeptide involved in the biosynthesis of an epothilone, wherein said nucleotide sequence is selected from the group consisting of: the complement of nucleotides 1900-3171 of SEQ ID NO:1, nucleotides 3415-5556 of SEQ ID NO:1, nucleotides 7610-11875 of SEQ ID NO:1, nucleotides 7643-8920 of SEQ ID NO:1, nucleotides 9236-10201 of SEQ ID NO:1, nucleotides 10529-11428 of SEQ ID NO:1, nucleotides 11549-11764 of SEQ ID NO:1, nucleotides 11872-16104 of SEQ ID NO:1, nucleotides 12085-12114 of SEQ ID NO:1, nucleotides 12223-12246 of SEQ ID NO:1, nucleotides 12466-12507 of SEQ ID NO:1, nucleotides 12928-12960 of SEQ ID NO:1, nucleotides 13516-13566 of SEQ ID NO:1, nucleotides 13633-13680 of SEQ ID NO:1, nucleotides 13876-13923 of SEQ ID NO:1, nucleotides 14313-14334 of SEQ ID NO:1, nucleotides 14473-14547 of SEQ ID NO:1, nucleotides 14578-14607 of SEQ ID NO:1, nucleotides 14623-14692 of SEQ ID NO:1, nucleotides 15673-15693 of SEQ ID NO:1, nucleotides 15724-15762 of SEQ ID NO:1, nucleotides 14788-15639 of SEQ ID NO:1, nucleotides 15901-15924 of SEQ ID NO:1, nucleotides 16251-21749 of SEQ ID NO:1, nucleotides 16269-17546 of SEQ ID NO:1, nucleotides 17865-18827 of SEQ ID NO:1, nucleotides 18855-19361 of SEQ ID NO:1, nucleotides 20565-21302 of SEQ ID NO:1, nucleotides 21414-21626 of SEQ ID NO:1, nucleotides 21746-43519 of SEQ ID NO:1, nucleotides 21860-23116 of SEQ ID NO:1, nucleotides 23431-24397 of SEQ ID NO:1, nucleotides 25184-25942 of SEQ ID NO:1, nucleotides 26045-26263 of SEQ ID NO:1, nucleotides 26318-27595 of SEQ ID NO:1, nucleotides 27911-28876 of SEQ ID NO:1, nucleotides 29678-30429 of SEQ ID NO:1, nucleotides 30539-30759 of SEQ ID NO:1, nucleotides 30815-32092 of SEQ ID NO:1, nucleotides 32408-33373 of SEQ ID NO:1, nucleotides 33401-33889 of SEQ ID NO:1, nucleotides 35042-35902 of SEQ ID NO:1, nucleotides 35930-36667 of SEQ ID NO:1, nucleotides 36773-36991 of SEQ ID NO:1, nucleotides 37052-38320 of SEQ ID NO:1, nucleotides 38636-39598 of SEQ ID NO:1, nucleotides 39635-40141 of SEQ ID NO:1, nucleotides 41369-42256 of SEQ ID NO:1, nucleotides 42314-43048 of SEQ ID NO:1, nucleotides 43163-43378 of SEQ ID NO:1, nucleotides 43524-54920 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, nucleotides 51534-52657 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, nucleotides 54540-54758 of SEQ ID NO:1, nucleotides 54935-62254 of SEQ ID NO:1, nucleotides 55028-56284 of SEQ ID NO:1, nucleotides 56600-57565 of SEQ ID) NO:1, nucleotides 57593-58087 of SEQ ID NO:1, nucleotides 59366-60304 of SEQ ID) NO:1, nucleotides 60362-61099 of SEQ ID NO:1, nucleotides 61211-61426 of SEQ ID NO:1, nucleotides 61427-62254 of SEQ ID NO:1, nucleotides 62369-63628 of SEQ ID NO:1, nucleotides 67334-68251 of SEQ ID NO:1, and nucleotides 1-68750 SEQ ID NO:1.

Brief Summary Text (18):
In yet another preferred embodiment, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes at least one polypeptide involved in the biosynthesis of an epothilone, wherein said nucleotide sequence comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence selected from the group consisting of: the complement of nucleotides 1900-3171 of SEQ ID NO:1, nucleotides 3415-5556 of SEQ ID NO:1, nucleotides 7610-11875 of SEQ ID NO:1, nucleotides 7643-8920 of SEQ ID NO:1, nucleotides 9236-10201 of SEQ ID NO:1, nucleotides 10529-11428 of SEQ ID NO:1, nucleotides 11549-11764 of SEQ ID NO:1, nucleotides 11872-16104 of SEQ ID NO:1, nucleotides 12085-12114 of SEQ ID NO:1, nucleotides 12223-12246 of SEQ ID NO:1, nucleotides 12466-12507 of SEQ ID NO:1, nucleotides 12928-12960 of SEQ ID NO:1, nucleotides 13516-13566 of SEQ ID NO:1, nucleotides 13633-13680 of SEQ ID NO:1, nucleotides 13876-13923 of SEQ ID NO:1, nucleotides 14313-14334 of SEQ ID NO:1, nucleotides 14473-14547 of SEQ ID NO:1, nucleotides 14578-14607 of SEQ ID NO:1, nucleotides 14623-14692 of SEQ ID NO:1, nucleotides 15673-15693 of SEQ ID NO:1, nucleotides 15724-15762 of SEQ ID NO:1, nucleotides 14788-15639 of SEQ ID NO:1, nucleotides 15901-15924 of SEQ ID NO:1, nucleotides 16251-21749 of SEQ ID NO:1, nucleotides 16269-17546 of SEQ ID NO:1, nucleotides 17865-18827 of SEQ ID NO:1, nucleotides 18855-19361 of SEQ ID NO:1, nucleotides 20565-21302 of SEQ ID NO:1, nucleotides 21414-21626 of SEQ ID NO:1, nucleotides 21746-43519 of SEQ ID NO:1, nucleotides 21860-23116 of SEQ ID NO:1, nucleotides 23431-24397 of SEQ ID NO:1, nucleotides 25184-25942 of SEQ ID NO:1, nucleotides 26045-26263 of SEQ ID NO:1, nucleotides 26318-27595 of SEQ ID NO:1, nucleotides 27911-28876 of SEQ ID NO:1, nucleotides 29678-30429 of SEQ ID NO:1, nucleotides 30539-30759 of SEQ ID NO:1, nucleotides 30815-32092 of SEQ ID NO:1, nucleotides 32408-33373 of SEQ ID NO:1, nucleotides 33401-33889 of SEQ ID NO:1, nucleotides 35042-35902 of SEQ ID NO:1, nucleotides 35930-36667 of SEQ ID NO:1, nucleotides 36773-36991 of SEQ ID NO:1, nucleotides 37052-38320 of SEQ ID NO:1, nucleotides 38636-39598 of SEQ ID NO:1, nucleotides 39635-40141 of SEQ ID NO:1, nucleotides 41369-42256 of SEQ ID NO:1, nucleotides 42314-43048 of SEQ ID NO:1, nucleotides 43163-43378 of SEQ ID NO:1, nucleotides 43524-54920 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, nucleotides 51534-52657 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, nucleotides 54540-54758 of SEQ ID NO:1, nucleotides 54935-62254 of SEQ ID NO:1, nucleotides 55028-56284 of SEQ ID NO:1, nucleotides 56600-57565 of SEQ ID NO:1, nucleotides 57593-58087 of SEQ ID NO:1, nucleotides 59366-60304 of SEQ ID NO:1, nucleotides 60362-61099 of SEQ ID NO:1, nucleotides 61211-61426 of SEQ ID NO:1,

nucleotides 61427-62254 of SEQ ID NO:1, nucleotides 62369-63628 of SEQ ID NO:1, nucleotides 67334-68251 of SEQ ID NO:1, and nucleotides 1-68750 SEQ ID NO:1.

Brief Summary Text (19):
The present invention also provides a chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid molecule of the invention. Further, the present invention provides a recombinant vector comprising such a chimeric gene, wherein the vector is capable of being stably transformed into a host cell. Still further, the present invention provides a recombinant host cell comprising such a chimeric gene, wherein the host cell is capable of expressing the nucleotide sequence that encodes at least one polypetide necessary for the biosynthesis of an epothilone. In a preferred embodiment, the recombinant host cell is a bacterium belonging to the order Actinomycetales, and in a more preferred embodiment the recombinant host cell is a strain of Streptomyces. In other embodiments, the recombinant host cell is any other bacterium amenable to fermentation, such as a pseudomonad or E. coli. Even further, the present invention provides a Bac clone comprising a nucleic acid molecule of the invention, preferably Bac clone pEPO15.

Brief Summary Text (20):
In another aspect, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes an epothilone synthase domain.

Brief Summary Text (21):
According to one embodiment, the epothilone synthase domain is a .beta.-ketoacyl-synthase (KS) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 11-437 of SEQ ID NO:2, amino acids 7-432 of SEQ ID NO:4, amino acids 39-457 of SEQ ID NO:5, amino acids 1524-1950 of SEQ ID NO:5, amino acids 3024-3449 of SEQ ID NO:5, amino acids 5103-5525 of SEQ ID NO:5, amino acids 35-454 of SEQ ID NO:6, amino acids 1522-1946 of SEQ ID NO: 6, and amino acids 32-450 of SEQ ID NO:7. According to this embodiment, said KS domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 11-437 of SEQ ID NO:2, amino acids 7-432 of SEQ ID NO:4, amino acids 39-457 of SEQ ID NO:5, amino acids 1524-1950 of SEQ ID NO:5, amino acids 3024-3449 of SEQ ID NO:5, amino acids 5103-5525 of SEQ ID NO:5, amino acids 35-454 of SEQ ID NO:6, amino acids 1522-1946 of SEQ ID NO: 6, and amino acids 32-450 of SEQ ID NO:7. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 7643-8920 of SEQ ID NO:1, nucleotides 16269-17546 of SEQ ID NO:1, nucleotides 21860-23116 of SEQ ID NO:1, nucleotides 26318-27595 of SEQ ID NO:1, nucleotides 30815-32092 of SEQ ID NO:1, nucleotides 37052-38320 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, and nucleotides 55028-56284 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence selected from the group consisting of: nucleotides 7643-8920 of SEQ ID NO:1, nucleotides 16269-17546 of SEQ ID NO:1, nucleotides 21860-23116 of SEQ ID NO:1, nucleotides 26318-27595 of SEQ ID NO:1, nucleotides 30815-32092 of SEQ ID NO:1, nucleotides 37052-38320 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, and nucleotides 55028-56284 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is selected from the group consisting of: nucleotides 7643-8920 of SEQ ID NO:1, nucleotides 16269-17546 of SEQ ID NO:1, nucleotides 21860-23116 of SEQ ID NO:1, nucleotides 26318-27595 of SEQ ID NO:1, nucleotides 30815-32092 of SEQ ID NO:1, nucleotides 37052-38320 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, and nucleotides 55028-56284 of SEQ ID NO:1.

Brief Summary Text (22):
According to another embodiment, the epothilone synthase domain is an acyltransferase (AT) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 543-864 of SEQ ID NO:2, amino acids 539-859 of SEQ ID NO:4, amino acids 563-884 of SEQ ID NO:5, amino acids 2056-2377 of SEQ ID NO:5, amino acids 3555-3876 of SEQ ID NO:5, amino acids 5631-5951 of SEQ ID NO:5, amino acids 561-881 of SEQ ID NO:6, amino acids 2053-2373 of SEQ ID NO:6, and amino acids 556-877 of SEQ ID NO:7. According to this embodiment, said AT domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 543-864 of SEQ ID NO:2, amino acids 539-859 of SEQ ID NO:4, amino acids 563-884 of SEQ ID NO:5, amino acids 2056-2377 of SEQ ID NO:5, amino acids 3555-3876 of SEQ ID NO:5, amino acids 5631-5951 of SEQ ID NO:5, amino acids 561-881 of SEQ ID NO:6, amino acids 2053-2373 of SEQ ID NO:6, and amino acids 556-877 of SEQ ID NO:7. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 9236-10201 of SEQ ID NO:1, nucleotides 17865-18827 of SEQ ID NO:1, nucleotides 23431-24397 of SEQ ID NO:1, nucleotides 27911-28876 of SEQ ID NO:1, nucleotides 32408-33373 of SEQ ID NO:1, nucleotides 38636-39598 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, and nucleotides 56600-57565 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence

selected from the group consisting of: nucleotides 9236-10201 of SEQ ID NO:1, nucleotides 17865-18827 of SEQ ID NO:1, nucleotides 23431-24397 of SEQ ID NO:1, nucleotides 27911-28876 of SEQ ID NO:1, nucleotides 32408-33373 of SEQ ID NO:1, nucleotides 38636-39598 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, and nucleotides 56600-57565 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is selected from the group consisting of: nucleotides 9236-10201 of SEQ ID NO:1, nucleotides 17865-18827 of SEQ ID NO:1, nucleotides 23431-24397 of SEQ ID NO:1, nucleotides 27911-28876 of SEQ ID NO:1, nucleotides 32408-33373 of SEQ ID NO:1, nucleotides 38636-39598 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, and nucleotides 56600-57565 of SEQ ID NO:1.

Brief Summary Text (23):
According to still another embodiment, the epothilone synthase domain is an enoyl reductase (ER) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 974-1273 of SEQ ID NO:2, amino acids 4433-4719 of SEQ ID NO:5, amino acids 6542-6837 of SEQ ID NO:5, and amino acids 1478-1790 of SEQ ID NO:7. According to this embodiment, said ER domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 974-1273 of SEQ ID NO:2, amino acids 4433-4719 of SEQ ID NO:5, amino acids 6542-6837 of SEQ ID NO:5, and amino acids 1478-1790 of SEQ ID NO:7. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 10529-11428 of SEQ ID NO:1, nucleotides 35042-35902 of SEQ ID NO:1, nucleotides 41369-42256 of SEQ ID NO:1, and nucleotides 59366-60304 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence selected from the group consisting of: nucleotides 10529-11428 of SEQ ID NO:1, nucleotides 35042-35902 of SEQ ID NO:1, nucleotides 41369-42256 of SEQ ID NO:1, and nucleotides 59366-60304 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is selected from the group consisting of: nucleotides 10529-11428 of SEQ ID NO:1, nucleotides 35042-35902 of SEQ ID NO:1, nucleotides 41369-42256 of SEQ ID NO:1, and nucleotides 59366-60304 of SEQ ID NO:1.

Brief Summary Text (24):
According to another embodiment, the epothilone synthase domain is an acyl carrier protein (ACP) domain, wherein said polypeptide comprises an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 1314-1385 of SEQ ID NO:2, amino acids 1722-1792 of SEQ ID NO:4, amino acids 1434-1506 of SEQ ID NO:5, amino acids 2932-3005 of SEQ ID NO:5, amino acids 5010-5082 of SEQ ID NO:5, amino acids 7140-7211 of SEQ ID NO:5, amino acids 1430-1503 of SEQ ID NO:6, amino acids 3673-3745 of SEQ ID NO:6, and amino acids 2093-2164 of SEQ ID NO:7. According to this embodiment, said ACP domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 1314-1385 of SEQ ID NO:2, amino acids 1722-1792 of SEQ ID NO:4, amino acids 1434-1506 of SEQ ID NO:5, amino acids 2932-3005 of SEQ ID NO:5, amino acids 5010-5082 of SEQ ID NO:5, amino acids 7140-7211 of SEQ ID NO:5, amino acids 1430-1503 of SEQ ID NO:6, amino acids 3673-3745 of SEQ ID NO:6, and amino acids 2093-2164 of SEQ ID NO:7. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 11549-11764 of SEQ ID NO:1, nucleotides 21414-21626 of SEQ ID NO:1, nucleotides 26045-26263 of SEQ ID NO:1, nucleotides 30539-30759 of SEQ ID NO:1, nucleotides 36773-36991 of SEQ ID NO:1, nucleotides 43163-43378 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 54540-54758 of SEQ ID NO:1, and nucleotides 61211-61426 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence selected from the group consisting of: nucleotides 11549-11764 of SEQ ID NO:1, nucleotides 21414-21626 of SEQ ID NO:1, nucleotides 26045-26263 of SEQ ID NO:1, nucleotides 30539-30759 of SEQ ID NO:1, nucleotides 36773-36991 of SEQ ID NO:1, nucleotides 43163-43378 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 54540-54758 of SEQ ID NO:1, and nucleotides 61211-61426 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is selected from the group consisting of: nucleotides 11549-11764 of SEQ ID NO:1, nucleotides 21414-21626 of SEQ ID NO:1, nucleotides 26045-26263 of SEQ ID NO:1, nucleotides 30539-30759 of SEQ ID NO:1, nucleotides 36773-36991 of SEQ ID NO:1, nucleotides 43163-43378 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 54540-54758 of SEQ ID NO:1, and nucleotides 61211-61426 of SEQ ID NO:1.

Brief Summary Text (25):
According to another embodiment, the epothilone synthase domain is a dehydratase (DH) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 869-1037 of SEQ ID NO:4, amino acids 3886-4048 of SEQ ID NO:5, amino acids 5964-6132 of SEQ ID NO:5, amino acids 2383-2551 of SEQ ID NO:6, and amino acids 887-1051 of SEQ ID NO:7. According to this embodiment, said DH domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 869-1037 of SEQ ID NO:4, amino acids 3886-4048 of SEQ ID NO:5, amino acids 5964-6132 of SEQ ID NO:5, amino

acids 2383-2551 of SEQ ID NO:6, and amino acids 887-1051 of SEQ ID NO:7. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 18855-19361 of SEQ ID NO:1, nucleotides 33401-33889 of SEQ ID NO:1, nucleotides 39635-40141 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, and nucleotides 57593-58087 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence selected from the group consisting of: nucleotides 18855-19361 of SEQ ID NO:1, nucleotides 33401-33889 of SEQ ID NO:1, nucleotides 39635-40141 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, and nucleotides 57593-58087 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is selected from the group consisting of: nucleotides 18855-19361 of SEQ ID NO:1, nucleotides 33401-33889 of SEQ ID NO:1, nucleotides 39635-40141 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, and nucleotides 57593-58087 of SEQ ID NO:1.

Brief Summary Text (26):
According to yet another embodiment, the epothilone synthase domain is a .beta.-ketoreductase reductase (KR) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 1439-1684 of SEQ ID NO:4, amino acids 1147-1399 of SEQ ID NO:5, amino acids 2645-2895 of SEQ ID NO:5, amino acids 4729-4974 of SEQ ID NO:5, amino acids 6857-7101 of SEQ ID NO:5, amino acids 1143-1393 of SEQ ID NO:6, amino acids 3392-3636 of SEQ ID NO:6, and amino acids 1810-2055 of SEQ ID NO:7. According to this embodiment, said KR domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 1439-1684 of SEQ ID NO:4, amino acids 1147-1399 of SEQ ID NO:5, amino acids 2645-2895 of SEQ ID NO:5, amino acids 4729-4974 of SEQ ID NO:5, amino acids 6857-7101 of SEQ ID NO:5, amino acids 1143-1393 of SEQ ID NO:6, amino acids 3392-3636 of SEQ ID NO:6, and amino acids 1810-2055 of SEQ ID NO:7. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 20565-21302 of SEQ ID NO:1, nucleotides 25184-25942 of SEQ ID NO:1, nucleotides 29678-30429 of SEQ ID NO:1, nucleotides 35930-36667 of SEQ ID NO:1, nucleotides 42314-43048 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, and nucleotides 60362-61099 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence selected from the group consisting of: nucleotides 20565-21302 of SEQ ID NO:1, nucleotides 25184-25942 of SEQ ID NO:1, nucleotides 29678-30429 of SEQ ID NO:1, nucleotides 35930-36667 of SEQ ID NO:1, nucleotides 42314-43048 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, and nucleotides 60362-61099 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is selected from the group consisting of: nucleotides 20565-21302 of SEQ ID NO:1, nucleotides 25184-25942 of SEQ ID NO:1, nucleotides 29678-30429 of SEQ ID NO:1, nucleotides 35930-36667 of SEQ ID NO:1, nucleotides 42314-43048 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, and nucleotides 60362-61099 of SEQ ID NO:1.

Brief Summary Text (27):
According to an additional embodiment, the epothilone synthase domain is a methyltransferase (MT) domain comprising an amino acid sequence substantially similar to amino acids 2671-3045 of SEQ ID NO:6. According to this embodiment, said MT domain preferably comprises amino acids 2671-3045 of SEQ ID NO:6. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to nucleotides 51534-52657 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of nucleotides 51534-52657 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is nucleotides 51534-52657 of SEQ ID NO:1.

Brief Summary Text (28):
According to another embodiment, the epothilone synthase domain is a thioesterase (TE) domain comprising an amino acid sequence substantially similar to amino acids 2165-2439 of SEQ ID NO:7. According to this embodiment, said TE domain preferably comprises amino acids 2165-2439 of SEQ ID NO:7. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to nucleotides 61427-62254 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of nucleotides 61427-62254 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is nucleotides 61427-62254 of SEQ ID NO:1.

Brief Summary Text (29):
In still another aspect, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a non-ribosomal peptide synthetase, wherein said

non-ribosomal peptide synthetase comprises an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: SEQ ID NO:3, amino acids 72-81 of SEQ ID NO:3, amino acids 118-125 of SEQ ID NO:3, amino acids 199-212 of SEQ ID NO:3, amino acids 353-363 of SEQ ID NO:3, amino acids 549-565 of SEQ ID NO:3, amino acids 588-603 of SEQ ID NO:3, amino acids 669-684 of SEQ ID NO:3, amino acids 815-821 of SEQ ID NO:3, amino acids 868-892 of SEQ ID NO:3, amino acids 903-912 of SEQ ID NO:3, amino acids 918-940 of SEQ ID NO:3, amino acids 1268-1274 of SEQ ID NO:3, amino acids 1285-1297 of SEQ ID NO:3, amino acids 97:3-1256 of SEQ ID NO:3, and amino acids 1344-1351 of SEQ ID NO:3. According to this embodiment, said non-ribosomal peptide synthetase preferably comprises an amino acid sequence selected from the group consisting of: SEQ ID NO:3, amino acids 72-81 of SEQ ID NO:3, amino acids 118-125 of SEQ ID NO:3, amino acids 199-212 of SEQ ID NO:3, amino acids 353-363 of SEQ ID NO:3, amino acids 549-565 of SEQ ID NO:3, amino acids 5838-603 of SEQ ID NO:3, amino acids 669-684 of SEQ ID NO:3, amino acids 815-821 of SEQ ID NO:3, amino acids 868-892 of SEQ ID NO:3, amino acids 903-912 of SEQ ID NO:3, amino acids 918-940 of SEQ ID NO:3, amino acids 1268-1274 of SEQ ID NO:3, amino acids 1285-1297 of SEQ ID NO:3, amino acids 973-1256 of SEQ ID NO:3, and amino acids 1344-1351 of SEQ ID NO:3. Also, according to this embodiment, said nucleotide sequence preferably is substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 11872-16104 of SEQ ID NO:1, nucleotides 12085-12114 of SEQ ID NO:1, nucleotides 12223-12246 of SEQ ID NO:1, nucleotides 12466-12507 of SEQ ID NO:1, nucleotides 12928-12960 of SEQ ID NO:1, nucleotides 13516-13566 of SEQ ID NO:1, nucleotides 13633-13680 of SEQ ID NO:1, nucleotides 13876-13923 of SEQ ID NO:1, nucleotides 14313-14334 of SEQ ID NO:1, nucleotides 14473-14547 of SEQ ID NO:1, nucleotides 14578-14607 of SEQ ID NO:1, nucleotides 14623-14692 of SEQ ID NO:1, nucleotides 15673-15693 of SEQ ID NO:1, nucleotides 15724-15762 of SEQ ID NO:1, nucleotides 14788-15639 of SEQ ID NO:1, and nucleotides 15901-15924 of SEQ ID NO:1. According to this embodiment, said nucleotide sequence more preferably comprises a consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair nucleotide portion identical in sequence to a respective consecutive 20, 25, 30, 35, 40, 45, or 50 (preferably 20) base pair portion of a nucleotide sequence selected from the group consisting of: nucleotides 11872-16104 of SEQ ID NO:1, nucleotides 12085-12114 of SEQ ID NO:1, nucleotides 12223-12246 of SEQ ID NO:1, nucleotides 12466-12507 of SEQ ID NO:1, nucleotides 12928-12960 of SEQ ID NO:1, nucleotides 13516-13566 of SEQ ID NO:1, nucleotides 13633-13680 of SEQ ID NO:1, nucleotides 13876-13923 of SEQ ID NO:1, nucleotides 14313-14334 of SEQ ID NO:1, nucleotides 14473-14547 of SEQ ID NO:1, nucleotides 14578-14607 of SEQ ID NO:1, nucleotides 14623-14692 of SEQ ID NO:1, nucleotides 15673-15693 of SEQ ID NO:1, nucleotides 15724-15762 of SEQ ID NO:1, nucleotides 14788-15639 of SEQ ID NO:1, and nucleotides 15901-15924 of SEQ ID NO:1. In addition, according to this embodiment, said nucleotide sequence most preferably is selected from the group consisting of: nucleotides 11872-16104 of SEQ ID NO:1, nucleotides 120835-12114 of SEQ ID NO:1, nucleotides 12223-12246 of SEQ ID NO:1, nucleotides 12466-12507 of SEQ ID NO:1, nucleotides 12928-12960 of SEQ ID NO:1, nucleotides 13516-13566 of SEQ ID NO:1, nucleotides 13633-13680 of SEQ ID NO:1, nucleotides 13876-13923 of SEQ ID NO:1, nucleotides 14313-14334 of SEQ ID NO:1, nucleotides 14473-14547 of SEQ ID NO:1, nucleotides 14578-14607 of SEQ ID NO:1, nucleotides 14623-14692 of SEQ ID NO:1, nucleotides 15673-15693 of SEQ ID NO:1, nucleotides 15724-15762 of SEQ ID NO:1, nucleotides 14788-15639 of SEQ ID NO:1, and nucleotides 15901-15924 of SEQ ID NO:1.

Brief Summary Text (30):
The present invention further provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs:2-23.

Brief Summary Text (31):
In accordance with another aspect, the present invention also provides methods for the recombinant production of polyketides such as epothilones in quantities large enough to enable their purification and use in pharmaceutical formulations such as those for the treatment of cancer. A specific advantage of these production methods is the chirality of the molecules produced; production in transgenic organisms avoids the generation of populations of racemic mixtures, within which some enantiomers may have reduced activity. In particular, the present invention provides a method for heterologous expression of epothilone in a recombinant host, comprising: (a) introducing into a host a chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid molecule of the invention that comprises a nucleotide sequence that encodes at least one polypeptide involved in the biosynthesis of epothilone; and (b) growing the host in conditions that allow biosynthesis of epothilone in the host. The present invention also provides a method for producing epothilone, comprising: (a) expressing epothilone in a recombinant host by the aforementioned method; and (b) extracting epothilone from the recombinant host.

Brief Summary Text (32):
According to still another aspect, the present invention provides an isolated polypeptide comprising an amino acid sequence that consists of an epothilone synthase domain.

Brief Summary Text (33):
According to one embodiment, the epothilone synthase domain is a .beta.-ketoacyl-synthase (KS) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 11-437 of SEQ ID NO:2, amino acids 7-432 of

SEQ ID NO:4, amino acids 39-457 of SEQ ID NO:5, amino acids 1524-1950 of SEQ ID NO:5, amino acids 3024-3449 of SEQ ID NO:5, amino acids 5103-5525 of SEQ ID NO:5, amino acids 35-454 of SEQ ID NO:6, amino acids 1522-1946 of SEQ ID NO: 6, and amino acids 32-450 of SEQ ID NO:7. According to this embodiment, said KS domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 11-437 of SEQ ID NO:2, amino acids 7-432 of SEQ ID NO:4, amino acids 39-457 of SEQ ID NO:5, amino acids 1524-1950 of SEQ ID NO:5, amino acids 3024-3449 of SEQ ID NO:5, amino acids 5103-5525 of SEQ ID NO:5, amino acids 35-454 of SEQ ID NO:6, amino acids 1522-1946 of SEQ ID NO: 6, and amino acids 32-450 of SEQ ID NO:7.

Brief Summary Text (34):
According to another embodiment, the epothilone synthase domain is an acyltransferase (AT) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 543-864 of SEQ ID NO:2, amino acids 539-859 of SEQ ID NO:4, amino acids 563-884 of SEQ ID NO:5, amino acids 2056-2377 of SEQ ID NO:5, amino acids 3555-3876 of SEQ ID NO:5, amino acids 5631-5951 of SEQ ID NO:5, amino acids 561-881 of SEQ ID NO:6, amino acids 2053-2373 of SEQ ID NO:6, and amino acids 556-877 of SEQ ID NO:7. According to this embodiment, said AT domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 543-864 of SEQ ID NO:2, amino acids 539-859 of SEQ ID NO:4, amino acids 563-884 of SEQ ID NO:5, amino acids 2056-2377 of SEQ ID NO:5, amino acids 3555-3876 of SEQ ID NO:5, amino acids 5631-5951 of SEQ ID NO:5, amino acids 561-881 of SEQ ID NO:6, amino acids 2053-2373 of SEQ ID NO:6, and amino acids 556-877 of SEQ ID NO:7.

Brief Summary Text (35):
According to still another embodiment, the epothilone synthase domain is an enoyl reductase (ER) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 974-1273 of SEQ ID NO:2, amino acids 4433-4719 of SEQ ID NO:5, amino acids 6542-6837 of SEQ ID NO:5, and amino acids 1478-1790 of SEQ ID NO:7. According to this embodiment, said ER domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 974-1273 of SEQ ID NO:2, amino acids 4433-4719 of SEQ ID NO:5, amino acids 6542-6837 of SEQ ID NO:5, and amino acids 1478-1790 of SEQ ID NO:7.

Brief Summary Text (36):
According to another embodiment, the epothilone synthase domain is an acyl carrier protein (ACP) domain, wherein said polypeptide comprises an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 1314-1385 of SEQ ID NO:2, amino acids 1722-1792 of SEQ ID NO:4, amino acids 1434-1506 of SEQ ID NO:5, amino acids 2932-3005 of SEQ ID NO:5, amino acids 5010-5082 of SEQ ID NO:5, amino acids 7140-7211 of SEQ ID NO:5, amino acids 1430-1503 of SEQ ID NO:6, amino acids 3673-3745 of SEQ ID NO:6, and amino acids 2093-2164 of SEQ ID NO:7. According to this embodiment, said ACP domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 1314-1385 of SEQ ID NO:2, amino acids 1722-1792 of SEQ ID NO:4, amino acids 1434-1506 of SEQ ID NO:5, amino acids 2932-3005 of SEQ ID NO:5, amino acids 5010-5082 of SEQ ID NO:5, amino acids 7140-7211 of SEQ ID NO:5, amino acids 1430-1503 of SEQ ID NO:6, amino acids 3673-3745 of SEQ ID NO:6, and amino acids 2093-2164 of SEQ ID NO:7.

Brief Summary Text (37):
According to another embodiment, the epothilone synthase domain is a dehydratase (DH) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 869-1037 of SEQ ID NO:4, amino acids 3886-4048 of SEQ ID NO:5, amino acids 5964-6132 of SEQ ID NO:5, amino acids 2383-2551 of SEQ ID NO:6, and amino acids 887-1051 of SEQ ID NO:7. According to this embodiment, said DH domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 869-1037 of SEQ ID NO:4, amino acids 3886-4048 of SEQ ID NO:5, amino acids 5964-6132 of SEQ ID NO:5, amino acids 2383-2551 of SEQ ID NO:6, and amino acids 887-1051 of SEQ ID NO:7.

Brief Summary Text (38):
According to yet another embodiment, the epothilone synthase domain is a .beta.-ketoreductase (KR) domain comprising an amino acid sequence substantially similar to an amino acid sequence selected from the group consisting of: amino acids 1439-1684 of SEQ ID NO:4, amino acids 1147-1399 of SEQ ID NO:5, amino acids 2645-2895 of SEQ ID NO:5, amino acids 4729-4974 of SEQ ID NO:5, amino acids 6857-7101 of SEQ ID NO:5, amino acids 1143-1393 of SEQ ID NO:6, amino acids 3392-3636 of SEQ ID NO:6, and amino acids 1810-2055 of SEQ ID NO:7. According to this embodiment, said KR domain preferably comprises an amino acid sequence selected from the group consisting of: amino acids 1439-1684 of SEQ ID NO:4, amino acids 1147-1399 of SEQ ID NO:5, amino acids 2645-2895 of SEQ ID NO:5, amino acids 4729-4974 of SEQ ID NO:5, amino acids 6857-7101 of SEQ ID NO:5, amino acids 1143-1393 of SEQ ID NO:6, amino acids 3392-3636 of SEQ ID NO:6, and amino acids 1810-2055 of SEQ ID NO:7.

Brief Summary Text (39):
According to an additional embodiment, the epothilone synthase domain is a methyl-transferase (MT) domain comprising an amino acid sequence substantially similar to amino acids 2671-3045 of SEQ ID NO:6. According to this embodiment, said MT domain preferably comprises amino acids 2671-3045 of SEQ ID NO:6.

Brief Summary Text (40):
According to another embodiment, the epothilone synthase domain is a thioesterase (TE) domain comprising an amino acid sequence substantially similar to amino acids 2165-2439 of SEQ ID NO:7. According to this embodiment, said TE domain preferably comprises amino acids 2165-2439 of SEQ ID NO:7.

Brief Summary Text (44):
Associated With/Operatively Linked: Refers to two DNA sequences that are related physically or functionally. For example, a promoter or regulatory DNA sequence is said to be "associated with" a DNA sequence that codes for an RNA or a protein if the two sequences are operatively linked, or situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence.

Brief Summary Text (45):
Chimeric Gene: A recombinant DNA sequence in which a promoter or regulatory DNA sequence is operatively linked to, or associated with, a DNA sequence that codes for an mRNA or which is expressed as a protein, such that the regulator DNA sequence is able to regulate transcription or expression of the associated DNA sequence. The regulator DNA sequence of the chimeric gene is not normally operatively linked to the associated DNA sequence as found in nature.

Brief Summary Text (46):
Coding DNA Sequence: A DNA sequence that is translated in an organism to produce a protein.

Brief Summary Text (47):
Domain: That part of a polyketide synthase necessary for a given distinct activity. Examples include acyl carrier protein (ACP), .beta.-ketosynthase (KS), acyltransferase (AT), .beta.-ketoreductase (KR), dehydratase (DH), enoylreductase (ER), and thioesterase (TE) domains.

Brief Summary Text (48):
Epothilones: 16-membered macrocyclic polyketides naturally produced by the bacterium Sorangium cellulosum strain So ce90, which mimic the biological effects of taxol. In this application, "epothilone" refers to the class of polyketides that includes epothilone A and epothilone B, as well as analogs thereof such as those described in WO 98/25929.

Brief Summary Text (49):
Epothilone Synthase: A polyketide synthase responsible for the biosynthesis of epothilone.

Brief Summary Text (50):
Gene: A defined region that is located within a genome and that, besides the aforementioned coding DNA sequence, comprises other, primarily regulatory, DNA sequences responsible for the control of the expression, that is to say the transcription and translation, of the coding portion.

Brief Summary Text (51):
Heterologous DNA Sequence: A DNA sequence not naturally associated with a host cell into which it is introduced, including non-naturally occurring multiple copies of a naturally occurring DNA sequence.

Brief Summary Text (52):
Homologous DNA Sequence: A DNA sequence naturally associated with a host cell into which it is introduced.

Brief Summary Text (53):
Homologous Recombination: Reciprocal exchange of DNA fragments between homologous DNA molecules.

Brief Summary Text (54):
Isolated: In the context of the present invention, an isolated nucleic acid molecule or an isolated enzyme is a nucleic acid molecule or enzyme that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated nucleic acid molecule or enzyme may exist in a purified form or may exist in a non-native environment such as, for example, a recombinant host cell.

Brief Summary Text (57):
NRPS gene: One or more genes encoding NRPSs for producing functional secondary metabolites, e.g., epothilones A and B, when under the direction of one or more compatible control elements.

Brief Summary Text (58):
Nucleic Acid Molecule: A linear segment of single- or double-stranded DNA or RNA that can be isolated from any source. In the context of the present invention, the nucleic acid molecule is preferably a segment of DNA.

Brief Summary Text (60):
PKS: A polyketide synthase, which is a complex of enzymatic activities (domains) responsible for the biosynthesis of polyketides including, for example, ketoreductase, dehydrates, acyl carrier protein, enoylreductase, ketoacyl ACP synthase, and acyltransferase. A functional PKS is one that catalyzes the synthesis of a polyketide.

Brief Summary Text (61):
PKS Genes: One or more genes encoding various polypeptides required for producing functional polyketides, e.g., epothilones A and B, when under the direction of one or more compatible control elements.

Brief Summary Text (62):
Substantially Similar: With respect to nucleic acids, a nucleic acid molecule that has at least 60 percent sequence identity with a reference nucleic acid molecule. In a preferred embodiment, a substantially similar DNA sequence is at least 80% identical to a reference DNA sequence; in a more preferred embodiment, a substantially similar DNA sequence is at least 90% identical to a reference DNA sequence; and in a most preferred embodiment, a substantially similar DNA sequence is at least 95% identical to a reference DNA sequence. A substantially similar DNA sequence preferably encodes a protein or peptide having substantially the same activity as the protein or peptide encoded by the reference DNA sequence. A substantially similar nucleotide sequence typically hybridizes to a reference nucleic acid molecule, or fragments thereof, under the following conditions: hybridization at 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO.sub.4 pH 7.0, 1 mM EDTA at 50.degree. C.; wash with 2.times.SSC, 1% SDS, at 50.degree. C. With respect to proteins or peptides, a substantially similar amino acid sequence is an amino acid sequence that is at least 90% identical to the amino acid sequence of a reference protein or peptide and has substantially the same activity as the reference protein or peptide.

Brief Summary Text (63):
Transformation: A process for introducing heterologous nucleic acid into a host cell or organism.

Brief Summary Text (64):
Transformed/Transgenic/Recombinant: Refers to a host organism such as a bacterium into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A "non-transformed", "non-transgenic", or "non-recombinant" host refers to a wild-type organism, i.e., a bacterium, which does not contain the heterologous nucleic acid molecule.

Brief Summary Text (67):
SEQ ID NO:1 is the nucleotide sequence of a 68750 bp contig containing 22 open reading frames (ORFs), which comprises the epothilone biosynthesis genes.

Brief Summary Text (68):
SEQ ID NO:2 is the protein sequence of a type I polyketide synthase (EPOS A) encoded by epoA (nucleotides 7610-11875 of SEQ ID NO:1).

Brief Summary Text (70):
SEQ ID NO:4 is the protein sequence of a type I polyketide synthase (EPOS B) encoded by epoB (nucleotides 16251-21749 of SEQ ID NO:1).

Brief Summary Text (71):
SEQ ID NO:5 is the protein sequence of a type I polyketide synthase (EPOS C) encoded by epoC (nucleotides 21746-43519 of SEQ ID NO:1).

Brief Summary Text (72):
SEQ ID NO:6 is the protein sequence of a type I polyketide synthase (EPOS D) encoded by epoD (nucleotides 43524-54920 of SEQ ID NO:1).

Brief Summary Text (73):
SEQ ID NO:7 is the protein sequence of a type I polyketide synthase (EPOS E) encoded by epoE (nucleotides 54935-62254 of SEQ ID NO:1).

Brief Summary Text (100):
The genes involved in the biosynthesis of epothilones can be isolated using the techniques according to the present invention. The preferable procedure for the isolation of epothilone biosynthesis genes requires the isolation of genomic DNA from an organism identified as producing epothilones A and B, and the transfer of the isolated DNA on a suitable plasmid or vector to a host organism that does not normally produce the polyketide, followed by the identification of transformed host colonies to which the epothilone-producing ability has been conferred. Using a technique such as .lambda.::Tn5 transposon mutagenesis (de Bruijn & Lupski,

Gene 27: 131-149 (1984)), the exact region of the transforming epothilone-conferring DNA can be more precisely defined. Alternatively or additionally, the transforming epothilone-conferring DNA can be cleaved into smaller fragments and the smallest that maintains the epothilone-conferring ability further characterized. Whereas the host organism lacking the ability to produce epothilone may be a different species from the organism from which the polyketide derives, a variation of this technique involves the transformation of host DNA into the same host that has had its epothilone-producing ability disrupted by mutagenesis. In this method, an epothilone-producing organism is mutated and non-epothilone-producing mutants are isolated. These are then complemented by genomic DNA isolated from the epothilone-producing parent strain.

Brief Summary Text (101):
A further example of a technique that can be used to isolate genes required for epothilone biosynthesis is the use of transposon mutagenesis to generate mutants of an epothilone-producing organism that, after mutagenesis, fails to produce the polyketide. Thus, the region of the host genome responsible for epothilone production is tagged by the transposon and can be recovered and used as a probe to isolate the native genes from the parent strain. PKS genes that are required for the synthesis of polyketides and that are similar to known PKS genes may be isolated by virtue of their sequence homology to the biosynthetic genes for which the sequence is known, such as those for the biosynthesis of rifamycin or soraphen. Techniques suitable for isolation by homology include standard library screening by DNA hybridization.

Brief Summary Text (102):
Preferred for use as a probe molecule is a DNA fragment that is obtainable from a gene or another DNA sequence that plays a part in the synthesis of a known polyketide. A preferred probe molecule comprises a 1.2 kb SmaI DNA fragment encoding the ketosynthase domain of the fourth module of the soraphen PKS (U.S. Pat. No. 5,716,849), and a more preferred probe molecule comprises the .beta.-ketoacyl synthase domains from the first and second modules of the rifamycin PKS (Schupp et al., FEMS Microbiology Letters 159: 201-207 (1998)). These can be used to probe a gene library of an epothilone-producing microorganism to isolate the PKS genes responsible for epothilone biosynthesis.

Brief Summary Text (103):
Despite the well-known difficulties with PKS gene isolation in general and despite the difficulties expected to be encountered with the isolation of epothilone biosynthesis genes in particular, by using the methods described in the instant specification, biosynthetic genes for epothilones A and B can surprisingly be cloned from a microorganism that produces that polyketide. Using the methods of gene manipulation and recombinant production described in this specification, the cloned PKS genes can be modified and expressed in transgenic host organisms.

Brief Summary Text (104):
The isolated epothilone biosynthetic genes can be expressed in heterologous hosts to enable the production of the polyketide with greater efficiency than might be possible from native hosts. Techniques for these genetic manipulations are specific for the different available hosts and are known in the art. For example, heterologous genes can be expressed in Streptomyces and other actinomycetes using techniques such as those described in McDaniel et al., Science 262: 1546-1550 (1993) and Kao et al., Science 265: 509-512 (1994), both of which are incorporated herein by reference. See also, Rowe et al., Gene 216: 215-223 (1998); Holmes et al., EMBO Journal 12(8): 3183-3191 (1993) and Bibb et al., Gene 38: 215-226 (1985), all of which are incorporated herein by reference.

Brief Summary Text (105):
Alternately, genes responsible for polyketide biosynthesis, i.e., epothilone biosynthetic genes, can also be expressed in other host organisms such as pseudomonads and E. coli. Techniques for these genetic manipulations are specific for the different available hosts and are known in the art. For example, PKS genes have been sucessfully expressed in E. coli using the pT7-7 vector, which uses the T7 promoter. See, Tabor et al., Proc. Natl. Acad. Sci. USA 82: 1074-1078 (1985), incorporated herein by reference. In addition, the expression vectors pKK223-3 and pKK223-2 can be used to express heterologous genes in E. coli, either in transcriptional or translational fusion, behind the tac or trc promoter. For the expression of operons encoding multiple ORFs, the simplest procedure is to insert the operon into a vector such as pKK223-3 in transcriptional fusion, allowing the cognate ribosome binding site of the heterologous genes to be used. Techniques for overexpression in gram-positive species such as Bacillus are also known in the art and can be used in the context of this invention (Quax et al., in: Industrial Microorganisms: Basic and Applied Molecular Genetics, Eds. Baltz et al., American Society for Microbiology, Washington (1993)).

Brief Summary Text (106):
Other expression systems that may be used with the epothilone biosynthetic genes of the invention include yeast and baculovirus expression systems. See, for example, "The Expression of Recombinant Proteins in Yeasts," Sudbery, P. E., Curr. Opin. Biotechnol. 7(5): 517-524 (1996); "Methods for Expressing Recombinant Proteins in Yeast," Mackay, et al., Editor(s): Carey, Paul R., Protein Eng. Des. 105-153, Publisher: Academic, San Diego, Calif. (1996);

"Expression of heterologous gene products in yeast," Pichuantes, et al., Editor(s): Cleland, J. L., Craik, C. S., Protein Eng. 129-161, Publisher: Wiley-Liss, New York, N.Y. (1996); WO 98/27203; Kealey et al., Proc. Natl. Acad. Sci. USA 95: 505-509 (1998); "Insect Cell Culture: Recent Advances, Bioengineering Challenges And Implications In Protein Production," Palomares, et al., Editor(s): Galindo, Enrique; Ramirez, Octavio T., Adv. Bioprocess Eng. Vol. II, Invited Pap. Int. Symp., 2nd (1998) 25-52, Publisher: Kluwer, Dordrecht, Neth; "Baculovirus Expression Vectors," Jarvis, Donald L., Editor(s): Miller, Lois K., Baculoviruses 389-431, Publisher: Plenum, New York, N. Y. (1997); "Production Of Heterologous Proteins Using The Baculovirus/Insect Expression System," Grittiths, et al., Methods Mol. Biol. (Totowa, N.J.) 75 (Basic Cell Culture Protocols (2nd Edition)) 427-440 (1997); and "Insect Cell Expression Technology," Luckow, Verne A., Protein Eng. 183-218, Publisher: Wiley-Liss, New York, N.Y. (1996); all of which are incorporated herein by reference.

Brief Summary Text (107):
Another consideration for expression of PKS genes in heterologous hosts is the requirement of enzymes for posttranslational modification of PKS enzymes by phosphopantetheinylation before they can synthesize polyketides. However, the enzymes responsible for this modification of type I PKS enzymes, phosphopantetheinyl (P-pant) transferases are not normally present in many hosts such as E. coli. This problem can be solved by coexpression of a P-pant transferase with the PKS genes in the heterologous host, as described by Kealey et al., Proc. Natl. Acad. Sci. USA 95: 505-509 (1998), incorporated herein by reference.

Brief Summary Text (108):
Therefore, for the purposes of polyketide production, the significant criteria in the choice of host organism are its ease of manipulation, rapidity of growth (i.e. fermentation), possession or the proper molecular machinery for processes such as posttranslational modification, and its lack of susceptibility to the polyketide being overproduced. Most preferred host organisms are actinomycetes such as strains of Streptomyces. Other preferred host organisms are pseudomonads and E. coli. The above-described methods of polyketide production have significant advantages over the technology currently used in the preparation of the compounds. These advantages include the cheaper cost of production, the ability to produce greater quantities of the compounds, and the ability to produce compounds of a preferred biological enantiomer, as opposed to racemic mixtures inevitably generated by organic synthesis. Compounds produced by heterologous hosts can be used in medical (e.g. cancer treatment in the case of epothilones) as well as agricultural applications.

Brief Summary Text (110):
The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified. Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Ausubel (ed.), Current Protocols in Molecular Biology, John Wiley and Sons, Inc. (1994); T. Maniatis, E. F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor laboratory, Cold Spring Harbor, N.Y., (1989); and by T. J. Silhavy, M. L. Berman, and L. W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984).

Detailed Description Text (2):
Cultivation of an Epothilone-Producing Strain of Sorangium cellulosum

Detailed Description Text (6):
To generate a Bac library, S. cellulosum cells cultivated as described in Example 1 above are embedded into agarose blocks, lysed, and the liberated genomic DNA is partially digested by the restriction enzyme HindIII. The digested DNA is separated on an agarose gel by pulsed-field electrophoresis. Large (approximately 90-150 kb) DNA fragments are isolated from the agarose gel and ligated into the vector pBelobacII. pBelobacII contains a gene encoding chloramphenicol resistance, a multiple cloning site in the lacZ gene providing for blue/white selection on appropriate medium, as well as the genes required for the replication and maintenance of the plasmid at one or two copies per cell. The ligation mixture is used to transform Escherichia coli DH10B electrocompetent cells using standard electroporation techniques. Chloramphenicol-resistant recombinant (white, lacZ mutant) colonies are transferred to a positively charged nylon membrane filter in 384 3.times.3 grid format. The clones are lysed and the DNA is cross-linked to the filters. The same clones are also preserved as liquid cultures at -80.degree. C.

Detailed Description Text (8):
Screening the Bac Library of Sorangium cellulosum 90 for the Presence of Type I Polyketide Synthase-Related Sequences

Detailed Description Text (9):
The Bac library filters are probed by standard Southern hybridization procedures. The DNA probes used encode .beta.-ketoacyl synthase domains from the first and second modules of the rifamycin polyketide synthase (Schupp et al., FEMS Microbiology Letters 159: 201-207 (1998)). The probe DNAs are generated by PCR with primers flanking each ketosynthase domain using the plasmid pNE95 as the template (pNE95 equals cosmid 2 described in Schupp et al. (1998)). 25 ng

of PCR-amplified DNA is isolated from a 0.5% agarose gel and labeled with .sup.32 P-dCTP using a random primer labeling kit (Gibco-BRL, Bethesda Md., U.S.A.) according to the manufacturer's instructions. Hybridization is at 65.degree. C. for 36 hours and membranes are washed at high stringency (3 times with 0.1.times.SSC and 0.5% SDS for 20 min at 65.degree. C.). The labeled blot is exposed on a phosphorescent screen and the signals are detected on a Phospholmager 445SI (screen and 445SI from Molecular Dynamics). This results in strong hybridization of certain Bac clones to the probes. These clones are selected and cultured overnight in 5 mls of Luria broth (LB) at 37.degree. C. Bac DNA from the Bac clones of interest is isolated by a typical miniprep procedure. The cells are resuspended in 200 .mu.l lysozyme solution (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, 5 mg/ml lysozyme), lysed in 400 .mu.l lysis solution (0.2 N NaOH and 2% SDS), the proteins are precipitated (3.0 M potassium acetate, adjusted to pH5.2 with acetic acid), and the Bac DNA is precipitated with isopropanol. The DNA is resuspended in 20 .mu.l of nuclease-free distilled water, restricted with BamHI (New England Biolabs, Inc.) and separated on a 0.7% agarose gel. The gel is blotted by Southern hybridization as described above and probed under conditions described above, with a 1.2 kb SmaI DNA fragment encoding the ketosynthase domain of the fourth module of the soraphen polyketide synthase as the probe (see, U.S. Pat. No. 5,716,849). Five different hybridization patterns are observed. One clone representing each of the five patterns is selected and named pEPO15, pEPO20, pEPO30, pEPO31, and pEPO33, respectively.

Detailed Description Text (12):
The DNA of the five selected Bac clones is digested with BamHI and random fragments are subcloned into pBluescript II SK+ (Stratagene) at the BamHI site. Subclones carrying inserts between 2 and 10 kb in size are selected for sequencing of the flanking ends of the inserts and also probed with the 1.2 SmaI probe as described above. Subclones that show a high degree of sequence homology to known polyketide synthases and/or strong hybridization to the soraphen ketosynthase domain are used for gene disruption experiments.

Detailed Description Text (17):
Gene Disruptions in Sorangium cellulosum BCE28/2 Using the Subcloned BamHI Fragments

Detailed Description Text (19):
Integration of the pCIB132-derived plasmids into the chromosome of Sorangium cellulosum BCE28/2 by homologous recombination is verified by Southern hybridization. For this experiment, complete DNA from 5-10 tranconjugants per transferred BamHI fragment is isolated (from 10 ml cultures grown in medium G52-H for three days) applying the method described by Pospiech and Neumann, Trends Genet. 11: 217 (1995). For the Southern blot, the DNA isolated as described above is cleaved either with the restriction enzymes BglII, ClaI, or NotI, and the respective BamHI inserts or pCIB132 are used as 32P labelled probes.

Detailed Description Text (21):
Analysis of the Effect of the Integrated BamHI Fragments on Epothilone Production by Sorangium cellulosum After Gene Disruption

Detailed Description Text (23):
Quantitative determination of the epothilone produced takes place after incubation of the cultures at 30.degree. C. and 180 rpm for 7 days. The complete culture broth is filtered by suction through a 150 .mu.m nylon filter. The resin remaining on the filter is then resuspended in 10 ml isopropanol and extracted by shaking the suspension at 180 rpm for 1 hour. 1 ml is removed from this suspension and centrifuged at 12,000 rpm in an Eppendorff Microfuge. The amount of epothilones A and B therein is determined by means of an HPLC and detection at 250 nm with a UV_DAD detector (HPLC with Waters-Symetry C18 column and a gradient of 0.02% phosphoric acid 60%-0% and acetonitril 40%-100%).

Detailed Description Text (24):
Transconjugants with three different integrated BamHI fragments subcloned from pEPO15, namely transconjugants with the BamHI fragment of plasmid pEPO15-21, transconjugants with the BamHI fragment of plasmid pEPO15-4-5, and transconjugants with the BamHI fragment of plasmid pEPO15-4-1, are tested in the manner described above. HPLC analysis reveals that all transconjugants no longer produce epothilone A or B. By contrast, epothilone A and B are detectable in a concentration of 2-4 mg/l in transconjugants with BamHI fragments integrated that are derived from pEPO20, pEPO30, pEPO31, pEPO33, and in the parental strain BCE28/2.

Detailed Description Text (28):
Plasmid DNA is isolated from the strain Escherichia coli DH10B [pEPO15-21], and the nucleotide sequence of the 2.3-kb BamHI insert in pEPO15-21 is determined. Automated DNA sequencing is done on the double-stranded DNA template by the dideoxynucleotide chain termination method, using Applied Biosystems model 377 sequencers. The primers used are the universal reverse primer (5'GGA AAC AGC TAT GAC CAT G 3' (SEQ ID NO:24)) and the universal forward primer (5'GTA AAA CGA CGG CCA GT 3' (SEQ ID NO:25)). In subsequent rounds of sequencing reactions, custom-synthesized oligonucleotides, designed for the 3' ends of the previously determined sequences, are used to extend and join contigs. Both strands are entirely sequenced, and every nucleotide is sequenced at least two times. The nucleotide sequence is compiled using the program Sequencher vers. 3.0 (Gene Codes Corporation), and analyzed using the University of

Wisconsin Genetics Computer Group programs. The nucleotide sequence of the 2213-bp insert corresponds to nucleotides 20779-22991 of SEQ ID NO:1.

Detailed Description Text (30):
Plasmid DNA is isolated from the strain Escherichia coli DH10B [pEPO15-4-1], and the nucleotide sequence of the 3.9-kb BamHI insert in pEPO15-4-1 is determined as described in (A) above. The nucleotide sequence of the 3909-bp insert corresponds to nucleotides 16876-20784 of SEQ ID NO:1.

Detailed Description Text (32):
Plasmid DNA is isolated from the strain Escherichia coli DH10B [pEPO15-4-5], and the nucleotide sequence of the 2.3-kb BamHI insert in pEPO15-4-5 is determined as described in (A) above. The nucleotide sequence of the 2233-bp insert corresponds to nucleotides 42528-44760 of SEQ ID NO:1.

Detailed Description Text (34):
Subcloning and Ordering of DNA Fragments from pEPO15 Containing Epothilone Biosynthesis Genes

Detailed Description Text (36):
The BamHI insert of pEPO15-21 is isolated and DIG-labeled (Non-radioactive DNA labeling and detection system, Boehringer Mannheim), and used as a probe in DNA hybridization experiments at high stringency against pEPO15-NH1, pEPO15-NH2, pEPO15-NH6, pEPO15-NH24, pEPO15-H2.7 and pEPO15-H3.0. Strong hybridization signal is detected for pEPO15-NH24, indicating that pEPO15-21 is contained within pEPO15-NH24.

Detailed Description Text (37):
The BamHI insert of pEPO15-4-1 is isolated and DIG-labeled as above, and used as a probe in DNA hybridization experiments at high stringency against pEPO15-NH1, pEPO15-NH2, pEPO15-NH6, pEPO15-NH24, pEPO15-H2.7 and pEPO15-H3.0. Strong hybridization signals are detected for pEPO15-NH24 and pEPO15-H2.7. Nucleotide sequence data generated from one end each of pEPO15-NH24 and pEPO15-H2.7 are also in complete agreement with the previously determined sequence of the BamHI insert of pEPO15-4-1. These experiments demonstrate that pEPO15-4-1 (which contains one internal HindIII site) overlaps pEPO15-H2.7 and pEPO15-NH24, and that pEPO15-H2.7 and pEPO15-NH24, in this order, are contiguous.

Detailed Description Text (38):
The BamHI insert of pEPO15-4-5 is isolated and DIG-labeled as above, and used as a probe in DNA hybridization experiments at high stringency against pEPO15-NH1, pEPO15-NH2, pEPO15-NH6, pEPO15-NH24, pEPO15-H2.7 and pEPO15-H3.0. Strong hybridization signal is detected for pEPO15-NH2, indicating that pEPO15-21 is contained within pEPO15-NH2.

Detailed Description Text (39):
Nucleotide sequence data is generated from both ends of pEPO15-NH2 and from the end of pEPO15-NH24 that does not overlap with pEPO15-4-1. PCR primers NH24 end "B": GTGACTGGCGCCTGGAATCTGCATGAGC (SEQ ID NO:26), NH2 end "A": AGCGGGAGCTTGCTAGACATTCTGTTTC (SEQ ID NO:27), and NH2 end "B": GACGCGCCTCGGGCAGCGCCCCAA (SEQ ID NO:28), pointing towards the HindIII sites, are designed based on these sequences and used in amplification reactions with pEPO15 and, in separate experiments, with Sorangium cellulosum So ce90 genomic DNA as the templates. Specific amplification is found with primer pair NH24 end "B" and NH2 end "A" with both templates. The amplimers are cloned into pBluescript II SK- and completely sequenced. The sequences of the amplimers are identical, and also agree completely with the end sequences of pEPO15-NH24 and pEPO15-NH2, fused at the HindIII site, establishing that the HindIII fragments of pEPO15-NH2 and pEPO15-NH24 are, in this order, contiguous.

Detailed Description Text (40):
The HindIII insert of pEPO15-H2.7 is isolated and DIG-labeled as above, and used as a probe in a DNA hybridization experiment at high stringency against pEPO15 digested by NotI. A NotI fragment of about 9 kb in size shows a strong a hybridization, and is further subcloned into pBluescript II SK- that has been digested with NotI and dephosphorylated with calf intestinal alkaline phosphatase, to yield pEPO15-N9-16. The NotI insert of pEPO15-N9-16 is isolated and DIG-labeled as above, and used as a probe in DNA hybridization experiments at high stringency against pEPO15-NH1, pEPO15-NH2, pEPO15-NH6, pEPO15-NH24, pEPO15-H2.7 and pEPO15-H3.0. Strong hybridization signals are detected for pEPO15-NH6, and also for the expected clones pEPO15-H2.7 and pEPO15-NH24. Nucleotide sequence data is generated from both ends of pEPO15-NH6 and from the end of pEPO15-H2.7 that does not overlap with pEPO15-4-1. PCR primers are designed pointing towards the HindIII sites and used in amplification reactions with pEPO15 and, in separate experiments, with Sorangium cellulosum So ce90 genomic DNA as the templates. Specific amplification is found with primer pair pEPO15-NH6 end "B": CACCGAAGCGTCGATCTGGTCCATC (SEQ ID NO:29) and pEPO15-H2.7 end "A": CGGTCAGATCGACGACGGGCTTTCC (SEQ ID NO:30) with both templates. The amplimers are cloned into pBluescript II SK- and completely sequenced. The sequences of the amplimers are identical, and also agree completely with the end sequences of pEPO15-NH6 and pEPO15-H2.7, fused at the HindIII site, establishing that the HindIII fragments of pEPO15-NH6 and pEPO15-H2.7 are, in this order, contiguous.

Detailed Description Text (43):
Further Extension of the Subclone Contig Covering the Epothilone Biosynthesis Genes

Detailed Description Text (44):
An approximately 2.2 kb BamHI-HindIII fragment derived from the downstream end of the insert of pEPO15-NH2 and thus representing the downstream end of the subclone contig described in Example 9 is isolated, DIG-labeled, and used in Southern hybridization experiments against pEPO15 and pEPO15-NH2 DNAs digested with several enzymes. The strongly hybridizing bands are always found to be the same in size between the two target DNAs indicating that the Sorangium cellulosum So ce90 genomic DNA fragment cloned into pEPO15 ends with the HindIII site at the downstream end of pEPO15-NH2.

Detailed Description Text (45):
A cosmid DNA library of Sorangium cellulosum So ce90 is generated, using established procedures, in pScosTriplex-II (Ji, et al., Genomics 31: 185-192 (1996)). Briefly, high-molecular weight genomic DNA of Sorangium cellulosum So ce90 is partially digested with the restriction enzyme Sau3AI to provide fragments with average sizes of about 40 kb, and ligated to BamHI and Xbal digested pScosTriplex-II. The ligation mix is packaged with Gigapack III XL (Stratagene) and used to transfect E. coli XL1 Blue MR cells.

Detailed Description Text (46):
The cosmid library is screened with the approximately 2.2 kb BamHI-HindIII fragment, derived from the downstream end of the insert of pEPO15-NH2, used as a probe in colony hybridization. A strongly hybridizing clone, named pEPO4E7 is selected. pEPO4E7 DNA is isolated, digested with several restriction endonucleases, and probed in Southern hybridization experiments with the 2.2 kb BamHI-HindIII fragment. A strongly hybridizing NotI fragment of approximately 9 kb in size is selected and subcloned into pBluescript II SK- to yield pEPO4E7-N9-8. Further Southern hybridization experiments reveal that the approximately 9 kb NotI insert of pEPO4E7-N9-8 overlaps pEPO15-NH2 over 6 kb in a NotI-HindIII fragment, while the remaining approximately 3 kb HindIII- NotI fragment would extend the subclone contig described in Example 9. End sequencing reveals, however, that the downstream end of the insert of pEPO4E7-N9-8 contains the BamHI-NotI polylinker of pScosTriplex-II, thereby indicating that the genomic DNA insert of pEPO4E7 ends at a Sau3AI site within the extending HindIII-NotI fragment and that the NotI site is derived from pScosTriplex-II.

Detailed Description Text (51):
Nucleotide Sequence Determination of the Subclone Contig Covering the Epothilone Biosynthesis Genes

Detailed Description Text (53):
pEPO15-H2.7. Plasmid DNA is isolated from the strain Escherichia coli DH10B [pEPO15-H2.7], and the nucleotide sequence of the 2.7-kb BamHI insert in pEPO15-H2.7 is determined. Automated DNA sequencing is done on the double-stranded DNA template by the dideoxynucleotide chain termination method, using Applied Biosystems model 377 sequencers. The primers used are the universal reverse primer (5'GGA AAC AGC TAT GAC CAT G 3' (SEQ ID NO:24)) and the universal forward primer (5'GTA AAA CGA CGG CCA GT 3' (SEQ ID NO:25)). In subsequent rounds of sequencing reactions, custom-synthesized oligonucleotides, designed for the 3' ends of the previously determined sequences, are used to extend and join contigs.

Detailed Description Text (54):
pEPO15-NH6, pEPO15-NH24 and pEPO15-NH2. The HindIII inserts of these plasmids are isolated, and subjected to random fragmentation using a Hydroshear apparatus (Genomic Instrumentation Services, Inc.) to yield an average fragment size of 1-2 kb. The fragments are end-repaired using T4 DNA Polymerase and Klenow DNA Polymerase enzymes in the presence of desoxynucleotide triphosphates, and phosphorylated with T4 DNA Kinase in the presence of ribo-ATP. Fragments in the size range of 1.5-2.2 kb are isolated from agarose gels, and ligated into pBluescript II SK- that has been cut with EcoRV and dephosphorylated. Random subclones are sequenced using the universal reverse and the universal forward primers.

Detailed Description Text (59):
Nucleotide Sequence Analysis of the Epothilone Biosynthesis Genes

Detailed Description Text (61):
epoA (nucleotides 7610-11875 of SEQ ID NO:1) codes for EPOS A (SEQ ID NO:2), a type I polyketide synthase consisting of a single module, and harboring the following domains: .beta.-ketoacyl-synthase (KS) (nucleotides 7643-8920 of SEQ ID NO:1, amino acids 11-437 of SEQ ID NO:2); acyltransferase (AT) (nucleotides 9236-10201 of SEQ ID NO:1, amino acids 543-864 of SEQ ID NO:2); enoyl reductase (ER) (nucleotides 10529-11428 of SEQ ID NO:1, amino acids 974-1273 of SEQ ID NO:2); and acyl carrier protein homologous domain (ACP) (nucleotides 11549-11764 of SEQ ID NO:1, amino acids 1314-1385 of SEQ ID NO:2). Sequence comparisons and motif analysis (Haydock, et al. FEBS Lett. 374: 246-248 (1995); Tang, et al., Gene 216: 255-265 (1998)) reveal that the AT encoded by EPOS A is specific for malonyl-CoA. EPOS A should be involved in the initiation of epothilone biosynthesis by loading the acetate unit to the multienzyme complex that will eventually form part of the 2-methylthiazole ring (C26 and C20).

Detailed Description Text (69):
epoB (nucleotides 16251-21749 of SEQ ID NO:1) codes for EPOS B (SEQ ID NO:4), a type I polyketide synthase consisting of a single module, and harboring the following domains: KS (nucleotides 16269-17546 of SEQ ID NO:1, amino acids 7-432 of SEQ ID NO:4); AT (nucleotides 17865-18827 of SEQ ID NO:1, amino acids 539-859 of SEQ ID NO:4); dehydratase (DH) (nucleotides 18855-19361 of SEQ ID NO:1, amino acids 869-1037 of SEQ ID NO:4); .beta.-ketoreductase (KR) (nucleotides 20565-21302 of SEQ ID NO:1, amino acids 1439-1684 of SEQ ID NO:4); and ACP (nucleotides 21414-21626 of SEQ ID NO:1, amino acids 1722-1792 of SEQ ID NO:4). Sequence comparisons and motif analysis reveal that the AT encoded by EPOS B is specific for methylmalonyl-CoA. EPOS A should be involved in the first polyketide chain extension by catalysing the Claisen-like condensation of the 2-methyl-4-thiazolecarboxyl-S-PCP starter group with the methylmalonyl-S-ACp, and the concomitant reduction of the b-keto group of C17 to an enoyl.

Detailed Description Text (70):
epoC (nucleotides 21746-43519 of SEQ ID NO:1) codes for EPOS C (SEQ ID NO:5), a type I polyketide synthase consisting of 4 modules. The first module harbors a KS (nucleotides 21860-23116 of SEQ ID NO:1, amino acids 39-457 of SEQ ID NO:5); a malonyl CoA-specific AT (nucleotides 23431-24397 of SEQ ID NO:1, amino acids 563-884 of SEQ ID NO:5); a KR (nucleotides 25184-25942 of SEQ ID NO:1, amino acids 1147-1399 of SEQ ID NO:5); and an ACP (nucleotides 26045-26263 of SEQ ID NO:1, amino acids 1434-1506 of SEQ ID NO:5). This module incorporates an acetate extender unit (C14-C13) and reduces the .beta.-keto group at C15 to the hydroxyl group that takes part in the final lactonization of the epothilone macrolactone ring. The second module of EPOS C harbors a KS (nucleotides 26318-27595 of SEQ ID NO:1, amino acids 1524-1950 of SEQ ID NO:5); a malonyl CoA-specific AT (nucleotides 27911-28876 of SEQ ID NO:1, amino acids 2056-2377 of SEQ ID NO:5); a KR (nucleotides 29678-30429 of SEQ ID NO:1, amino acids 2645-2895 of SEQ ID NO:5); and an ACP (nucleotides 30539-30759 of SEQ ID NO:1, amino acids 2932-3005 of SEQ ID NO:5). This module incorporates an acetate extender unit (C12-C1 1) and reduces the .beta.-keto group at C13 to a hydroxyl group. Thus, the nascent polyketide chain of epothilone corresponds to epothilone A, and the incorporation of the methyl side chain at C12 in epothilone B would require a post-PKS C-methyltransferase activity. The formation of the epoxi ring at C13-C12 would also require a post-PKS oxidation step. The third module of EPOS C harbors a KS (nucleotides 30815-32092 of SEQ ID NO:1, amino acids 3024-3449 of SEQ ID NO:5); a malonyl CoA-specific AT (nucleotides 32408-33373 of SEQ ID NO:1, amino acids 3555-3876 of SEQ ID NO:5); a DH (nucleotides 33401-33889 of SEQ ID NO:1, amino acids 3886-4048 of SEQ ID NO:5); an ER (nucleotides 35042-35902 of SEQ ID NO:1, amino acids 4433-4719 of SEQ ID NO:5); a KR (nucleotides 35930-36667 of SEQ ID NO:1, amino acids 4729-4974 of SEQ ID NO:5); and an ACP (nucleotides 36773-36991 of SEQ ID NO:1, amino acids 5010-5082 of SEQ ID NO:5). This module incorporates an acetate extender unit (C10-C9) and fully reduces the .beta.-keto group at C11. The fourth module of EPOS C harbors a KS (nucleotides 37052-38320 of SEQ ID NO:1, amino acids 5103-5525 of SEQ ID NO:5); a methylmalonyl CoA-specific AT (nucleotides 38636-39598 of SEQ ID NO:1, amino acids 5631-5951 of SEQ ID NO:5); a DH (nucleotides 39635-40141 of SEQ ID NO:1, amino acids 5964-6132 of SEQ ID NO:5); an ER (nucleotides 41369-42256 of SEQ ID NO:1, amino acids 6542-6837 of SEQ ID NO:5); a KR (nucleotides 42314-43048 of SEQ ID NO:1, amino acids 6857-7101 of SEQ ID NO:5); and an ACP (nucleotides 43163-43378 of SEQ ID NO:1, amino acids 7140-7211 of SEQ ID NO:5). This module incorporates a propionate extender unit (C24 and C8-C7) and fully reduces the .beta.-keto group at C9.

Detailed Description Text (71):
epoD (nucleotides 43524-54920 of SEQ ID NO:1) codes for EPOS D (SEQ ID NO:6), a type I polyketide synthase consisting of 2 modules. The first module harbors a KS (nucleotides 43626-44885 of SEQ ID NO:1, amino acids 35-454 of SEQ ID NO:6); a methylmalonyl CoA-specific AT (nucleotides 45204-46166 of SEQ ID NO:1, amino acids 561-881 of SEQ ID NO:6); a KR (nucleotides 46950-47702 of SEQ ID NO:1, amino acids 1143-1393 of SEQ ID NO:6); and an ACP (nucleotides 47811-48032 of SEQ ID NO:1, amino acids 1430-1503 of SEQ ID NO:6). This module incorporates a propionate extender unit (C23 and C6-C5) and reduces the .beta.-keto group at C7 to a hydoxyl group. The second module harbors a KS (nucleotides 48087-49361 of SEQ ID NO:1, amino acids 1522-1946 of SEQ ID NO: 6); a methylmalonyl CoA-specific AT (nucleotides 49680-50642 of SEQ ID NO:1, amino acids 2053-2373 of SEQ ID NO:6); a DH (nucleotides 50670-51176 of SEQ ID NO:1, amino acids 2383-2551 of SEQ ID NO:6); a methyltransferase (MT, nucleotides 51534-52657 of SEQ ID NO:1, amino acids 2671-3045 of SEQ ID NO:6); a KR (nucleotides 53697-54431 of SEQ ID NO:1, amino acids 3392-3636 of SEQ ID NO:6); and an ACP (nucleotides 54540-54758 of SEQ ID NO:1, amino acids 3673-3745 of SEQ ID NO:6). This module incorporates a propionate extender unit (C21 or C22 and C4-C3) and reduces the .beta.-keto group at C5 to a hydoxyl group. This reduction is somewhat unexpected, since epothilones contain a keto group at C5. Discrepancies of this kind between the deduced reductive capabilities of PKS modules and the redox state of the corresponding positions in the final polyketide products have been, however, reported in the literature (see, for example, Schwecke, et al., Proc. Nat. Acad. Sci. USA 92: 7839-7843 (1995) and Schupp, et al., FEMS Microbiology Letters 159: 201-207 (1998)). An important feature of epothilones is the presence of gem-methyl side groups at C4 (C21 and C22). The second module of EPOS D is predicted to incorporate a propionate unit into the growing polyketide chain, providing one; methyl side chain at C4. This module also contains a methyltransferase domain integrated into the PKS between the DH and the KR domains, in an arrangement similar to the one

seen in the HMWP1 yersiniabactin synthase (Gehring, A. M., DeMoll, E., Fetherston, J. D., Mori, I., Mayhew, G. F., Blattner, F. R., Walsh, C. T., and Perry, R. D.: Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by Yersinia pestis. Chem. Biol. 5, 573-586,1998). This MT domain in EPOS D is proposed to be responsible for the incorporation of the second methyl side group (C21 or C22) at C4.

Detailed Description Text (72):
epoE (nucleotides 54935-62254 of SEQ ID NO:1) codes for EPOS E (SEQ ID NO:7), a type I polyketide synthase consisting of one module, harboring a KS (nucleotides 55028-56284 of SEQ ID NO:1, amino acids 32-450 of SEQ ID NO:7); a malonyl CoA-specific AT (nucleotides 56600-57565 of SEQ ID NO:1, amino acids 556-877 of SEQ ID NO:7); a DH (nucleotides 57593-58087 of SEQ ID NO:1, amino acids 887-1051 of SEQ ID NO:7); a probably nonfunctional ER (nucleotides 59366-60304 of SEQ ID NO:1, amino acids 1478-1790 of SEQ ID NO:7); a KR (nucleotides 60362-61099 of SEQ ID NO:1, amino acids 1810-2055 of SEQ ID NO:7); an ACP (nucleotides 61211-61426 of SEQ ID NO:1, amino acids 2093-2164 of SEQ ID NO:7); and a thioesterase (TE) (nucleotides 61427-62254 of SEQ ID NO:1, amino acids 2165-2439 of SEQ ID NO:7). The ER domain in this module harbors an active site motif with some highly unusual amino acid substitutions that probably render this domain inactive. The module incorporates an acetate extender unit (C2-C1), and reduces the .beta.-keto at C3 to an enoyl group. Epothilones contain a hydroxyl group at C3, so this reduction also appears to be excessive as discussed for the second module of EPOS D. The TE domain of EPOS E takes part in the release and cyclization of the grown polyketide chain via lactonization between the carboxyl group of C1 and the hydroxyl group of C15.

Detailed Description Text (73):
Five ORFs are detected upstream of epoA in the sequenced region. The partially sequenced orf1has no homologues in the sequence databanks. The deduced protein product (Orf 2, SEQ ID NO:10) of orf2 (nucleotides 3171-1900 on the reverse complement strand of SEQ ID NO:1) shows strong similarities to hypothetical ORFs from Mycobacterium and Streptomyces coelicolor, and more distant similarities to carboxypeptidases and DD-peptidases of different bacteria. The deduced protein product of orf3 (nucleotides 3415-5556 of SEQ ID NO:1), Orf 3 (SEQ ID NO:11), shows homologies to Na/H antiporters of different bacteria. Orf 3 might take part in the export of epothilones from the producer strain. orf4 and orf5 have no homologues in the sequence databanks.

Detailed Description Text (74):
Eleven ORFs are found downstream of epoE in the sequenced region. epoF (nucleotides 62369-63628 of SEQ ID NO:1) codes for EPOS F (SEQ ID NO:8), a deduced protein, with strong sequence similarities to cytochrome P450 oxygenases. EPOS F may take part in the adjustment of the redox state of the carbons C12, C5, and/or C3. The deduced protein product of orf4 (nucleotides 67334-68251 of SEQ ID NO:1), Orf 14 (SEQ ID NO:22) shows strong similarities to GI:3293544, a hypothetic protein with no proposed function from Streptomyces coelicolor, and also to GI:2654559, the human embrionic lung protein. It is also more distantly related to cation efflux system proteins like GI:2623026 from Methanobacterium thermoautotrophicum, so it might also take part in the export of epothilones from the producing cells. The remaining ORFs (orf6-orf13 and orf15) show no homologies to entries in the sequence databanks.

Detailed Description Text (76):
Recombinant Expression of Epothilone Biosynthesis Genes

Detailed Description Text (77):
Epothilone synthase genes according to the present invention are expressed in heterologous organisms for the purposes of epothilone production at greater quantities than can be accomplished by fermentation of Sorangium cellulosum. A preferable host for heterologous expression is Streptomyces, e.g. Streptomyces coelicolor, which natively produces the polyketide actinorhodin. Techniques for recombinant PKS gene expression in this host are described in McDaniel et al., Science 262: 1546-1550 (1993) and Kao et al., Science 265: 509-512 (1994). See also, Holmes et al., EMBO Journal 12(8): 3183-3191 (1993) and Bibb et al., Gene 38: 215-226 (1985), as well as U.S. Pat. Nos. 5,521,077, 5,672,491, and 5,712,146, which are incorporated herein by reference.

Detailed Description Text (78):
According to one method, the heterologous host strain is engineered to contain a chromosomal deletion of the actinorhodin (act) gene cluster. Expression plasmids containing the epothilone synthase genes of the invention are constructed by transferring DNA from a temperature-sensitive donor plasmid to a recipient shuttle vector in E. coli (McDaniel et al. (1993) and Kao et al. (1994)), such that the synthase genes are built-up by homologous recombination within the vector. Alternatively, the epothilone synthase gene cluster is introduced into the vector by restriction fragment ligation. Following selection, e.g. as described in Kao et al. (1994), DNA from the vector is introduced into the act-minus Streptomyces coelicolor strain according to protocols set forth in Hopwood et al., Genetic Manipulation of Streptomyces. A Laboratory Manual (John Innes Foundation, Norwich, United Kingdom, 1985), incorporated herein by reference. The recombinant Streptomyces strain is grown on R2YE medium (Hopwood et al. (1985)) and produces epothilones. Alternatively, the epothilone synthase genes according to the present invention are expressed in other host organisms such as

pseudomonads, Bacillus, yeast, insect cells and/or E. coli. PKS and NRPS genes are preferably expressed in E. coli using the pT7-7 vector, which uses the T7 promoter. See, Tabor et al., Proc. Natl. Acad. Sci. USA 82: 1074-1078 (1985). In another embodiment, the expression vectors pKK223-3 and pKK223-2 are used to express PKS and NRPS genes in E. coli, either in transcriptional or translational fusion, behind the tac or trc promoter. Expression of PKS and NRPS genes in heterologous hosts, which do not naturally have the phosphopantetheinyl (P-pant) transferases needed for post-translational modification of PKS enzymes, requires the coexpression in the host of a P-pant transferase, as described by Kealey et al, Proc. Natl. Acad. Sci. USA 95: 505-509 (1998).

Detailed Description Text (80):
Isolation of Epothilones from Producing Strains

Detailed Description Text (81):
Examples of cultivation, fermentation, and extraction procedures for polyketide isolation, which are useful for extracting epothilones from both native and recombinant hosts according to the present invention, are given in WO 93/10121, incorporated herein by reference, in Example 57 of U.S. Pat. No. 5,639,949, in Gerth et al., J. Antibiotics 49: 560-563 (1996), and in Swiss patent application Ser. No. 396/98, filed Feb. 19,1998, and U.S. patent application No. 09/248,910 (that discloses also preferred mutant strains of Sorangium cellulosum), both of which are incorporated herein by reference. The following are procedures that are useful for isolating epothilones from cultured Sorangium cellulosum strains, e.g., So ce90 , and may also be used for the isolation of epothilone from recombinant hosts.

Detailed Description Text (82):
A: Cultivation of Epothilone-producing Strains

Detailed Description Text (114):
B: Effect of the Addition of Cyclodextrin and Cyclodextrin Derivatives to the Epothilone Concentrations Attained

Detailed Description Text (127):
All the cyclodextrin derivatives tested here are obtainable from the company Fluka, Buchs, CH. The tests are carried out in 200 ml agitating flasks with 50 ml culture volume. As controls, flasks with adsorber resin Amberlite XAD-16 (Rohm & Haas, Frankfurt, Germany) and without any adsorber addition are used. After incubation for 5 days, the following epothilone titres can be determined by HPLC:

Detailed Description Text (128):
Few of the cyclodextrins tested (2,6-di-o-methyl-.beta.-cyclodextrin, methyl-.beta.-cyclodextrin) display no effect on epothilone production at the concentrations used. 1-2% 2-hydroxy-propyl-.beta.-cyclodextrin and .beta.-cyclodextrin increase epothilone production in compared with production using no cyclodextrins.

Detailed Description Text (137):
G: Working Up of the Epothilones: Isolation from a 500 Liter Main Culture

Detailed Description Text (138):
The volume of harvest from the 500 liter main culture of example 2D is 450 liters and is separated using a Wesffalia clarifying separator Type SA-20-06 (rpm=6500) into the liquid phase (centrifugate+rinsing water=650 liters) and solid phase (cells=ca. 15 kg). The main part of the epothilones are found in the centrifugate, The centrifuged cell pulp contains <15% of the determined epothilone portion and is not further processed. The 650 liter centrifugate is then placed in a 4000 liter stirring vessel, mixed with 10 liters of Amberlite XAD-16 (centrifugate:resin volume=65:1) and stirred. After a period of contact of ca. 2 hours, the resin is centrifuged away in a Heine overflow centrifuge (basket content 40 liters; rpm=2800). The resin is discharged from the centrifuge and washed with 10-15 liters of deionised water. Desorption is effected by stirring the resin twice, each time in portions with 30 liters of isopropanol in 30 liter glass stirring vessels for 30 minutes. Separation of the isopropanol phase from the resin takes place using a suction filter. The isopropanol is then removed from the combined isopropanol phases by adding 15-20 liters of waiter in a vacuum-operated circulating evaporator (Schmid-Verdampfer) and the resulting water phase of ca. 10 liters is extracted 3.times. each time with 10 liters of ethyl acetate. Extraction is effected in 30 liter glass stirring vessels. The ethyl acetate extract is concentrated to 3-5 liters in a vacuum-operated circulating evaporator (Schmid-Verdampfer) and afterwards concentrated to dryness in a rotary evaporator (Buchi type) under vacuum. The result is an ethyl acetate extract of 50.2 g. The ethyl acetate extract is dissolved in 500 ml of methanol, the insoluble portions filtered off using a folded filter, and the solution added to a 10 kg Sephadex LH 20 column (Pharmacia, Uppsala, Sweden) (column diameter 20 cm, filling level ca. 1.2 m). Elution is effected with methanol as eluant. Epothilone A and B is present predominantly in fractions 21-23 (at a fraction size of 1 liter). These fractions are concentrated to dryness in a, vacuum on a rotary evaporator (total weight 9.0 g). These Sephadex peak fractions (9.0 g) are thereafter dissolved in 92 ml of acetonitrile:-water:-methylene chloride=50:40:2, the solution filtered through a folded filter and added to a RP column (equipment Prepbar 200, Merck; 2.0 kg

LiChrospher RP-18 Merck, grain size 12 .mu.m, column diameter 10 cm, filling level 42 cm; Merck, Darmstadt, Germany). Elution is effected with acetonitrile:water=3:7 (flow rate=500 ml/min.; retention time of epothilone A=ca. 51-59 mins.; retention time of epothilone B=ca. 60-69 mins.). Fractionation is monitored with a UV detector at 250 nm. The fractions are concentrated to dryness under vacuum on a Buchi-Rotavapor rotary evaporator. The weight of the epothilone A peak fraction is 700 mg, and according to HPLC (external standard) it has a content of 75.1%. That of the epothilone B peak fraction is 1980 mg, and the content according to HPLC (external standard) is 86.6%. Finally, the epothilone A fraction (700 mg) is crystallised from 5 ml of ethyl acetate:toluene=2:3, and yields 170 mg of epothilone A pure crystallisate [content according to HLPC (% of area)=94.3%]. Crystallisation of the epothilone B fraction (1980 mg) is effected from 18 ml of methanol and yields 1440 mg of epothilone B pure crystallisate [content according to HPLC; (% of area)=99.2%]. m.p. (Epothilone B): e.g. 124-125.degree. C.; .sup.1 H-NMR data for Epothilone B: 500 MHz-NMR, solvent: DMSO-d6. Chemical displacement .delta.in ppm relative to TMS. s=singlet; d=doublet; m=multiplet

Detailed Description Text (140):
Medical Uses of Recombinantly Produced Epothilones

Detailed Description Text (141):
Pharmaceutical preparations or compositions comprising epothilones are used for example in the treatment of cancerous diseases, such as various human solid tumors. Such anticancer formulations comprise, for example, an active amount of an epothilone together with one or more organic or inorganic, liquid or solid, pharmaceutically suitable carrier materials. Such formulations are delivered, for example, enterally, nasally, rectally, orally, or parenterally, particularly intramuscularly or intravenously. The dosage of the active ingredient is dependent upon the weight, age, and physical and pharmacokinetical condition of the patient and is further dependent upon the method of delivery. Because epothilones mimic the biological effects of taxol, epothilones may be substituted for taxol in compositions and methods utilizing taxol in the treatment of cancer. See, for example, U.S. Pat. Nos. 5,496,804, 5,565,478, and 5,641,803, all of which are incorporated herein by reference.

Detailed Description Text (142):
For example, for treatments, epothilone B is supplied in individual 2 ml glass vials formulated as 1 mg/1 ml of clear, colorless intravenous concentrate. The substance is formulated in polyethylene glycol 300 (PEG 300) and diluted with 50 or 100 ml 0.9% Sodium Chloride Injection, USP, to achieve the desired final concentration of the drug for infusion. It is administered as a single 30-minute intravenous infusion every 21 days (treatment three-weekly) for six cycles, or as a single 30-minute intravenous infusion every 7 days (weekly treatment).

Detailed Description Paragraph Table (1):
TABLE 1 ORF Start codon Stop codon Homology of deduced protein Proposed function of deduced protein orf1 outside of 1826 sequenced range orf2* 3171 1900 Hypothetical protein SP: Q11037; DD-peptidase SP:P15555 orf3 3415 5556 Na/H antiporter PID: D1017724 Transport orf4* 5992 5612 orf5 6226 6675 epoA 7610 11875 Type I polyketide synthase Epothilone synthase: Thiazole ring formation epoP 11872 16104 Non-ribosomal peptide synthetase Epothilone synthase: Thiazole ring formation epoB 16251 21749 Type I polyketide synthase Epothilone synthase: Polyketide backbone formation epoC 21746 43519 Type I polyketide synthase Epothilone synthase: Polyketide backbone formation epoD 43524 54920 Type I polyketide synthase Epothilone synthase: Polyketide backbone formation epoE 54935 62254 Type I polyketide synthase Epothilone synthase: Polyketide backbone formation epoF 62369 63628 Cytochrome P450 Epothilone macrolactone oxidase orf6 63779 64333 orf7* 64290 63853 orf8 64363 64920 orf9* 64727 64287 orf10 65063 65767 orf11* 65874 65008 orf12* 66338 65871 orf13 66667 67137 orf14 67334 68251 Hypothetical protein GI:3293544; Transport Cation efflux system protein GI:2623026 orf15 68346 outside of sequenced range *On the reverse complementer strand. Numbering according to SEQ ID NO:1.

Detailed Description Paragraph Table (5):
TABLE 3 Progress of a 10 liter fermentation duration of culture [d] Epothilone A [mg/l] Epothilone B [mg/l] 0 0 0 1 0 0 2 0.5 0.3 3 1.8 2.5 4 3.0 5.1 5 3.7 5.9 6 3.6 5.7

Detailed Description Paragraph Table (6):
TABLE 4 Progress of a 100 liter fermentation duration of culture [d] Epothilone A [mg/l] Epothilone B [mg/l] 0 0 0 1 0 0 2 0.3 0 3 0.9 1.1 4 1.5 2.3 5 1.6 3.3 6 1.8 3.7 7 1.8 3.5

Detailed Description Paragraph Table (7):
TABLE 5 Progress of a 500 liter fermentation duration of culture [d] Epothilone A [mg/l] Epothilone B [mg/l] 0 0 0 1 0 0 2 0 0 3 0.6 0.6 4 1.7 2.2 5 3.1 4.5 6 3.1 5.1

Detailed Description Paragraph Table (8):
TABLE 6 Progress of a 10 liter fermentation without adsorber. duration of culture [d] Epothilone A [mg/l] Epothilone B [mg/l] 0 0 0 1 0 0 2 0 0 3 0 0 4 0.7 0.7 5 0.7 1.0 6 0.8 1.3

Detailed Description Paragraph Table (10):
SEQUENCE LISTING <100> GENERAL INFORMATION: <160> NUMBER OF SEQ ID NOS: 30 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 1 <211> LENGTH: 68750 <212> TYPE: DNA <213> ORGANISM:

Sorangium cellulosum <400> SEQUENCE: 1 aagcttcgct cgacgccctc ttcgcccgcg ccacctctgc ccgtgtgctc
gatgatggcc 60 acggccgggc cacggagcgg catgtgctcg ccgaggcgcg cgggatcgag gacctccgcg 120 ccctccgaga
gcacctccgc atccaggaag gggggccgtc ctttcactgc atgtgcctcg 180 gcgacctgac ggtggagctc ctcgcgcacg
accagcccct cgcgtccatc agcttccacc 240 atgcccgcag cctgaggcac cccgactgga cctcggacgc gatgctcgtc
gacggccccg 300 cgctcgtccg gtggctcgcc gcgcgcggcg cgccgggtcc cctccgcgag tacgaagagg 360 agcgcgagcg
agcccgaacc gcgcaggagg cgaggcgcct gtggctcgcg gccgcgccgc 420 cctgcttcgc gcccgatctg ccccgcttcg
aggacgacgc caacgggctg ccgctcggcc 480 cgatgtcgcc tgaagtcgcc gaggccgagc ggcgcctccg cgcctcgtac
gcgactcctg 540 agctcgcctg tgccgcgctg ctcgcctggc tcgggacggg cgcgggtccc tggtccggat 600 atcccgccta
cgagatgctg ccagagaatc tgctcctcgg gtttggcctc ccgaccgcga 660 tcgccgcggc ctccgcgccc ggcacatcgg
aggccgctct ccgcggcgca gcgcggctgt 720 tcgcctcctg ggaggtcgta tcgagcaaga agagccagct cggcaacatc
cccgaagccc 780 tgtgggagcg gctccggacg atcgtccgcg cgatgggcaa tgccgacaac ctctctcgct 840 tcgagcgcgc
cgaggcgatc gcggcggagg tgcgccgcct gcgcgcacag ccggcgccct 900 tcgcggcggg cgccggcctg gcggtcgctg
gggtctcctc gagcggccgg ctctcgggcc 960 tcgtgaccga cggagacgca ttgtactccg gcgacggcaa cgacatcgtc
atgttccaac 1020 ccggccggat ctcgccggtc gtgctgctcg ccggaaccga tcccttcttc gagctcgcac 1080
cgcccctcag ccagatgctc ttcgtcgcgc acgccaacgc gggcaccatc tccaaggtcc 1140 tgacggaagg cagccccctc
atcgtgatgg caagaaacca ggcgcgaccg atgagcctcg 1200 tccacgctcg cgggttcatg gcgtgggtca accaggccat
ggtgcccgac cccgagcggg 1260 gcgcgccctt cgtcgtccag cgctcgacca tcatggaatt cgagcacccc acgcctcgtt
1320 gtctccacga gcccgccggc agcgctttct ccctcgcctg cgacgaggag cacctctact 1380 ggtgcgagct
ttcggctggc cggctcgagc tatggcgcca cccgcaccac cgccccggcg 1440 ccccgagccg cttcgcgtac ctcggcgagc
accccattgc ggcgacctgg tacccctcgc 1500 tcaccctcaa tgcgacccac gtgctgtggg ccgaccctga tcgcagggcc
atcctcgggg 1560 tcgacaagcg caccggcgta gagcccatcg tcctcgcgga gacgcgccat cccccggcgc 1620
acgtcgtgtc cgaggaccgg gacatcttcg cgcttaccgg acagcccgac tcccgcgact 1680 ggcacgtcga gcacatccgc
tccggcgcct ccaccgtcgt ggccgactac cagcgccagc 1740 tatgggaccg ccctgacatg gtgctcaatc ggcgcggcct
cttcttcacg acgaacgacc 1800 gcatcctgac gctcgcccgc agctgacatc gctcgacgcc gggccgctca tcgagggcgc
1860 ccggaccgag ctggcgaccc gccgctggcg ggccgcagct catgccgatt cggtggcgac 1920 gtagacgctg
cgccagaaac gctcgagagc ccccgagaac aggaagccgg cggattgtgt 1980 catcacgatc ccgatcagct cgcggcccgg
atcattgatc caggacgtcc cgaacccgcc 2040 gtcccaccca tagcgcccgg gcacctccga gaccgcgtcc ggcgccgtga
ccacggccat 2100 cccataaccc cagccgtgcg tctcgaagaa gcccgggaaa aacgaggacg ccgccttctg 2160
ggccggcgtg aggtgatcgg ccgtcatctc gcgcaccgag gcggcgctca agagccgccg 2220 gccctcgtgc acaccgccgt
tcatgagcat gcgcgcgaac aggaggtagt cgtccaccgt 2280 cgacacgagc ccggcggcgc ccgaagggaa cgccggcggg
ctggcatagg cgctctcggc 2340 cccgtcgcga tccatgcgcg tcttctcccc cgtctgctcg tcggtgaagt aaccgcagcc
2400 cgcgaaccga gcgagcttgt ccgccgggac gtgaaagtcg gtgtcccgca tcccgagcgg 2460 cgcgaggatg
cgctcgcgca cgaacgcatc gaagccctgg tcggccgcgc gccccacgag 2520 caccccctgc accaggctcc ccgtgttgta
catccactgc gcccccggct gatgcatgag 2580 cggcagcgtc ccgagccgcc ggatccactc gtctggcccg tgcggcgtca
tcggcaccgg 2640 ctgcgcgttg acgagcccga gctcgtcgat ggcccgctgg atcggcgacg atgcgtcgaa 2700
cgagattccg aagcccatcg tgaacgtcat caggtcgcgc accgtgatcg gccgctccgc 2760 gggcaccgtc tcgtcgatcg
gaccatcgat gcgcgccagc accttccggt tcgcgagctc 2820 cggcaaccat cggtcgacgg gggagtcgag gtcgagcttg
ccttcctcga cgagcatcat 2880 caccgccgtc gcggtgaccg ccttcgtcat cgaggcgatc cggaagatcg tgtcccgccg
2940 catgggcgcg ctgccgccga gctcggtcac gcccaccgcg tccacgtgca cgtcgtcgcc 3000 gcgcgcgacc
agccagaccg ctcccggcat ctgccccgcc gccacctccg ccgccatcac 3060 ctcgcgcgcg ggcgccagcg cgccggcccc
cgcgtcctgc cctggctgcc cctcctcctc 3120 ggccccaccc aacgcgcacc ccggcgccgc cacgctgatc aaagctccca
taaactcccg 3180 ccttctcatg accgtcgatg cctctccgag cgggggcgcc tgcccctgcc gagagcactg 3240
actgcccgcg cccgaaaaaa tcatcggtgc cccgtcacga tcgccgccgg gcgtggctcc 3300 gcccggccgc ccgctcgggc
gcccgcccct ggacgagcaa agctcgcccg cccgcgctca 3360 gcacgccgct tgccatgtcc ggcctgcacc cacaccgagg
agccacccac cctgatgcac 3420 ggcctcaccg agcggcaggt cctgctctcg ctcgtcaccc tcgcgctcat cctcgtgacc
3480 gcgcgcgcct ccggcgagct cgcgcggcgg ctgcgccagc ccgaggtgct cggggagctc 3540 ttcggcggcg
tcgtgctggg cccctccgtc gtcggcgcgc tcgcgcccgg gttccatcga 3600 gccctcttcc aggagccggc ggtcggggtc
gtgctctcgg gcatctcctg gataggcgcg 3660 ctcctcctgc tgctgatggc gggcatcgag gtcgacgtgg gcatcctgcg
caaggaggcg 3720 cgcccgggg cgctctcggc gctcggcgcg atcgcgcccc cgctcgcggc gggcgccgcc 3780
ttctcggcgc tcgtgctcga tcggcccctt ccgagcggcc tcttcctcgg gatcgtgctc 3840 tcggtgacgg cggtcagcgt
gatcgcgaag gtgctgatcg agcgcgagtc gatgcgccgc 3900 agctatgcgc aggtgacgct cgcggcgggg gtggtcagcg
aggtcgctgc ctgggtgctc 3960 gtcgcgatga cgtcgtcgag ctacggcgcg tcgcccgcgc tggcggtcgc ccggagcgcg
4020 ctcctggcga gcggattctt gctgttcatg gtgctcgtcg ggcggcggct cacccacctc 4080 gcgatgcgct
gggtggccga cgcgacgcgc gtctccaagg gacaggtgtc gctcgtcctc 4140 gtcctcacgt tcctggccgc ggcgctgacg
cagcggctcg gcctgcaccc gctgctcggc 4200 gcgttcgcgc tcggcgtgct gctcaacagc gctcctcgca ccaaccgccc
tctcctcgac 4260 ggcgtgcaga cgctcgtggc gggcctcttc gcgcctgtgt tcttcgtcct cgcgggcatg 4320
cgcgtcgacg tgtcgcagct gcgcacgccg gcggcgtggg ggacggtcgc gttgctgctg 4380 gcgaccgcga cggcggcgaa
ggtcgtcccc gccgcgctcg gcgcgcggct cggcgggctc 4440 aggggcagcg aggcggcgct cgtggcggtg ggcctgaaca
tgaagggcgg cacggacctc 4500 atcgtcgcga tcgtcggcgt cgagctcggg ctcctctcca acgaggctta tacgatgtac
4560 gccgtcgtcg cgctggtcac ggtgaccgcc tcacccgcgc tcctcatctg gctcgagaaa 4620 agggcgcctc
cgacgcagga ggagtcggct cgcctcgagc gcgaggaggc cgcgaggcgc 4680 gcgtacatcc ccggggtcga gcggatcctc
gtcccgatcg tggcgcacgc cctgcccggg 4740 ttcgccacgg acatcgtgga gagcatcgtc gcctccaagc gaaagctcgg
cgagacggtc 4800 gacatcacgg agctctccgt ggagcagcag gcgcccggcc catcgcgcgc cgcgggggag 4860
gcgagccggg ggctcgcgag gctcggcgcg cgcctccgcg tcggcatctg gcggcaaagg 4920 cgcgagctgc gcggctcgat
ccaggcgatc ctgcgcgcct cgcgggatca cgatctgctc 4980 gtgatcggcg cgcgatcgcc ggcgcgcgcg cgcggaatgt
cgttcggtcg cctgcaggac 5040 gcgatcgtcc agcgggccga gtccaacgtg ctcgtcgtgg tgggcgaccc tccggcggcg
5100 gagcgcgcct ccgcgcggcg gatcctcgtc ccgatcatcg gcctcgagta ctccttcgcc 5160 gccgccgatc
tcgcggccca cgtggcgctg gcgtgggacg ccgagctcgt gctgctcagc 5220 agcgcgcaga ccgatccggg cgcggtcgtc
tggcgcgatc gcgagccatc ccgggtgcgc 5280 gcggtggcgc ggagcgtcgt cgacgaggcg gtcttccggg ggcgccggct
cggcgtgcgc 5340 gtctcgtcgc gcgtgcacgt gggcgcgcac ccgagcgacg agataacgcg ggagctcgcg 5400
cgcgccccgt acgatctgct cgtgctcgga tgctacgacc atgggccgct cggccggctc 5460 tacctcggca gcacggtcga
gtcggtggtg gtccggagcc gggtgccggt cgcgttgctc 5520 gtcgcgcatg gagggactcg agagcaggtg aggtgaggct
tccaccgcgc tcgcccgtga 5580 ggaagcgagc gcccggctct gccgacgatc gtcactcccg gtccgtgtag gcgatcgtgc
5640 tgagcagcgc gttctccgcc tgacgcgagt cgagccgggt atgctgcacg acgatggggg 5700 cgtccgattc

gatcacgctg gcatagtccg tatcgcgcgg gatcggctcg ggttcggtca 5760 gatcgttgaa ccggacgtgc cgggtgcgcc
tcgctggaac ggtcacccgg taaggcccgg 5820 cggggtcgcg gtcgctgaag taaacggtga tggcgacctg cgcgtcccgg
tccgacgcat 5880 tcaacaggca ggccgtctca tggctcgtca tctgcggctc aggtccgttg ctcccgcctg 5940
ggatgtagcc ctctgcgatt gcacagcgcg tccgcccgat cggcttgtcc atgtgtcctc 6000 cctcctggct cctctttggc
agcctccctc tgctgtccag gagcgatggc ctcttcgctc 6060 gacgcgctcg gggatccatg gctgaggatc ctcgccgagc
gctccctgcc gaccggcgcg 6120 ccgagcgccg acgggctttg aaagcgcgcg accggccagc ccggacgcgg gcccgagagg
6180 gacagtgggt ccgccgtgaa gcagagaggc gatcgaggtg gtgagatgaa acacgtcgac 6240 acgggccgac
gattcggccg ccggataggg cacacgctcg gtcttctcgc gagcatggcg 6300 ctcgccggct gcggcggtcc gagcgagaaa
accgtgcagg gcacgcggct cgcgcccggc 6360 gccgatgcgc gcgtcaccgc cgacgtcgac cccgacgccg cgaccacgcg
gctggcggtg 6420 gacgtcgttc acctctcgcc gcccgagcgg ctcgaggccg gcagcgagcg gttcgtcgtc 6480
tggcagcgtc cgagccccga gtccccgtgg cgacgggtcg gagtgctcga ctacaatgct 6540 gacagccgaa gaggcaagct
ggccgagacg accgtgccgt atgccaactt cgagctgctc 6600 atcaccgccg agaagcagag cagcctcag tcgccatcgt
ctgccgccgt catcgggccg 6660 acgtctgtcg ggtgacatcg cgctatcagc agcgctgagc ccgccagcag gccccagggc
6720 cctgcctcga tggccttccc catcaccgct gcgcactcct ccagcgacgg ccgcgcagcg 6780 acggccgcgt
ccaagcaacc gccgtgccgg cgcggctcca cgcgcgcgac aggcgagcgt 6840 cctggcgcgg cctgcgcatc gctggaagga
tcggcggagc atggatagag aatcgaggat 6900 cgcgatcttt gttgccatcg cagccaacgt ggcgatcgcg gcggtcaagt
tcatcgccgc 6960 cgccgtgacc ggcagctcgg cgaggcgttt gccgacttcg gcggcgtccc gcgcgtgctg 7020
ctctacgaca acctcaagag cgccgtcgtc gagcgccacg gcgacgcgat ccggttccac 7080 cccacgctgc tggctctgtc
ggcgcattac cgcttcgagc cgcgccccgt cgccgtcgcc 7140 cgcggcaacg agaagggccg cgtccagcgc gccatcacgg
cgtggacgac atggcgcgga 7200 aacgtcgtcg taaccgccca gcaatgtcat gggaatggcc ccttgaaatg gccccttgag
7260 ggggctggcc ggggtcgacg atatcgcgcg atctccccgt caattcccga tggtaaaaga 7320 aaaatttgtc
atagatcgta agctgtgata gtggtctgtc ttacgttgcg tcttccgcac 7380 ctcgagcgag ttctctcgga taactttcaa
tttttccgag gggggcttgg tctctggttc 7440 ctcaggaagc ctgatcggga cgagctaatt cccatccatt tttttgaggc
tctgctcaaa 7500 gggattagat cgagtgagac agttcttttg cagtgcgcga agaacctggg cctcgaccgg 7560
aggacgatcg acgtccgcga gcgggtcagc cgctgaggat gtgcccgtcg tggcggatcg 7620 tcccatcgag cgcgcagccg
aagatccgat tgcgatcgtc ggagcgagtt gccgtctgcc 7680 cggtggcgtg atcgatctga gcgggttctg gacgctcctc
gagggctcgc gcgacaccgt 7740 cgggcgagtc cccgccgaac gctgggatgc agcagcgtgg tttgatcccg accccgatgc
7800 cccggggaag acgcccgtta cgcgcgcatc tttcctgagc gacgtagcct gcttcgacgc 7860 ctccttcttc
ggcatctcgc ctcgcgaagc gctgcggatg gaccctgcac atcgactctt 7920 gctggaggtg tgctgggagg cgctggagaa
cgccgcgatc gctccatcgg cgctcgtcgg 7980 tacggaaacg ggagtgttca tcgggatcgg cccgtccgaa tatgaggccg
cgctgccgca 8040 agcgacggcg tccgcagaga tcgacgctca tggcgggctg gggacgatgc ccagcgtcgg 8100
agcgggccga atctcgtatg ccctcgggct gcgagggccg tgtgtcgcgg tggatacggc 8160 ctattcgtcc tcgctggtgg
ccgttcatct ggcctgtcag agcttgcgct ccggggaatg 8220 ctccacggcc ctggctggtg gggtatcgct gatgttgtcg
ccgagcaccc tcgtgtggct 8280 ctcgaagacc cgggcgctgg ccagggacgg tcgctgcaag gcattttcgg cggaggccga
8340 tgggttcgga cgaggcgaag ggtgcgccgt cgtggtcctc aagcggctca gtggagcccg 8400 cgcggacggc
gatcggatat tggcggtgat tcgaggatcc gcgatcaatc acgacggtgc 8460 gagcagcggt ctgaccgtgc cgaacgggag
ctcccaagaa atcgtgctga aacgggccct 8520 ggcggacgca ggctgcgccg cgtcttcggt gggttatgtc gaggcacacg
gcacgggcac 8580 gacgcttggt gaccccatcg aaatccaagc tctgaatgcg gtatacggcc tcgggcgaga 8640
tgtcgccacg ccgctgctga tcgggtcggt gaagaccaac cttggccatc ctgagtatgc 8700 gtcggggatc actgggctgc
tgaaggtcgt cttgtccctt cagcacgggc agattcctgc 8760 gcacctccac gcgcaggcgc tgaaccccgg gatctcatgg
ggtgatcttc ggctgaccgt 8820 cacgcgcgcc cggacaccgt ggccggactg gaatacgccg cgacgggcgg gggtgagctc
8880 gttcggcatg agcgggacca acgcgcacgt ggtgctggaa gaggcgccgg cggcgacgtg 8940 cacaccgccg
gcgccggagc gaccggcaga gctgctggtg ctgtcggcaa ggaccgcgtc 9000 agccctggat gcacaggcgg cgcggctgcg
cgaccatctg gagacctacc cttcgcagtg 9060 tctgggcgat gtggcgttca gtctggcgac gacgcgcagc gcgatggagc
accggctcgc 9120 ggtggcggcg acgtcgaggg aggggctgcg ggcagccctg gacgctgcgg cgcagggaca 9180
gacgtcgccc ggtgcggtgc gcagtatcgc cgattcctca cgcggcaagc tcgcctttct 9240 cttcaccgga caggggcgc
agacgctggg catgggccgt gggctgtacg atgtatggtc 9300 cgcgttccgc gaggcgttcg acctgtgcgt gaggctgttc
aaccaggagc tcgaccggcc 9360 gctccgcgag gtgatgtggg ccgaaccggc cagcgtcgac gccgcgctgc tcgaccagac
9420 agccttcacc cagccggcgc tgttcacctt cgaatatgcg ctcgccgcgc tgtggcggtc 9480 gtggggtgta
gagccggagt tggtcgccgg ccatagcatc ggtgagctgg tggctgcctg 9540 cgtggcgggc gtgttctcgc ttgaggacgc
ggtgttcctg gtggctgcgc gcgggcgcct 9600 gatgcaggcg ctgccggccg gcggggcgat ggtgtcgatc gaggcgccgg
aggccgatgt 9660 ggctgctgcg gtggcgccgc acgcagcgtc ggtgtcgatc gccgcggtca acgctccgga 9720
ccaggtggtc atcgcgggcg ccgggcaacc cgtgcatgcg atcgcggcgg cgatggccgc 9780 gcgcggggcg cgaaccaagg
cgctccacgt ctcgcatgcg ttccactcac cgctcatggc 9840 cccgatgctg gaggcgttcg ggcgtgtggc cgagtcggtg
agctaccggc ggccgtcgat 9900 cgtcctggtc agcaatctga gcgggaaggc ttgcacagac gaggtgagct cgccgggcta
9960 ttgggtgcgc cacgcgcgag aggtggtgcg cttcgcggat ggagtgaagg cgctgcacgc 10020 ggccggtgcg
ggcaccttcg tcgaggtcgg tccgaaatcg acgctgctcg gcctggtgcc 10080 tgcctgcatg ccggacgccc ggccggcgct
gctcgcatcg tcgcgcgctg ggcgtgacga 10140 gccggcgacc gtgctcgagg cgctcggcgg gctctgggcc gtcggtggcc
tggtctcctg 10200 ggccggcctc ttcccctcag gggggcggcg ggtgccgctg cccacgtacc cttggcagcg 10260
cgagcgctac tggatcgaca cgaaagccga cgacgcggcg cgtggcgacc gccgtgctcc 10320 gggagcgggt cacgacgagg
tcgaggaggg gggcgcggtg cgcggcggcg accggcgcag 10380 cgctcggctc gaccatccgc cgcccgagag cggacgccgg
gagaaggtcg aggccgccgg 10440 cgaccgtccg ttccggctcg agatcgatga gccaggcgtg cttgatcacc tcgtgcttcg
10500 ggtcacggag cggcgcgccc ctggtctggg cgaggtcgag atcgccgtcg acgcggcggg 10560 gctcagcttc
aatgatgtcc agctcgcgct gggcatggtg cccgacgacc tgccgggaaa 10620 gcccaaccct ccgctgctgc tcggaggcga
gtgcgccggg cgcatcgtcg ccgtgggcga 10680 gggcgtgaac ggcctcgtgg tgggccaacc ggtcatcgcc ctttcggcgg
gagcgtttgc 10740 tacccacgtc accacgtcgg ctgcgctggt gctgcctcgg cctcaggcgc tctcggcgat 10800
cgaggcggcc gccatgcccg tcgcgtacct gacggcatgg tacgcgctcg acagaatagc 10860 ccgccttcag ccgggggagc
gggtgctgat ccatgcggcg accggcgggg tcggtctcgc 10920 cgcggtgcag tgggcgcagc acgtgggagc cgaggtccat
gcgacggccg gcacgcccga 10980 gaaacgcgcc tacctggagt cgctgggcgt gcggtatgtg agcgattccc gctcggaccg
11040 gttcgtcgcc gacgtgcgcg cgtggacggg cggcgaggga gtagacgtcg tgctcaactc 11100 gctctcgggc
gagctgatcg acaagagttt caatctcctg cgatcgcacg gccggtttgt 11160 ggagctcggc aagcgcgact gttacgcgga
taaccagctc gggctgcggc cgttcctgcg 11220 caatctctcc ttctcgctgg tggatctccg ggggatgatg ctcgagcggc
cggcgcgggt 11280 ccgtgcgctc ttggaggagc tcctcggcct gatcgcggca ggcgtgttca cccctccccc 11340
catcgcgacg ctcccgatcg cccgtgtcgc cgatgcgttc cggagcatgg cgcaggcgca 11400 gcatcttggg aagctcgtac

tcacgctggg tgacccggag gtccagatcc gtattccaac 11460 ccacgcaggc gccggcccgt ccaccgggga tcgggacctg ctcgacaggc tcgcgtcagc 11520 tgcgccggcc gcgcgcgcgg cggcgctgga ggcgttcctc cgtacgcagg tctcgcaggt 11580 gctgcgcacg cccgaaatca aggtcggcgc ggaggcgctg ttcacccgcc tcggcatgga 11640 ctcgctcatg gccgtggagc tgcgcaatcg tatcgaggcg agcctcaagc tgaagctgtc 11700 gacgacgttc ctgtccacgt cccccaatat cgccttgttg gcccaaaacc tgttggatgc 11760 tctcgccaca gctctctcct tggagcgggt ggcggcggag aacctacggg caggcgtgca 11820 aaacgacttc gtctcatcgg gcgcagatca agactgggaa atcattgccc tatgacgatc 11880 aatcagcttc tgaacgagct cgagcaccag ggtatcaagc tggcggccga tggggagcgc 11940 ctccagatac aggccccaa gaacgccctg aacccgaacc tgctcgctcg aatctccgag 12000 cacaaaagca cgatcctgac gatgctccgt cagagactcc ccgcagaatc catcgtgccc 12060 gccccagccg agcggcacgc tccgtttcct ctcacagaca tccaagaatc ctactggctg 12120 ggccggacag gagcgtttac ggtccccagc gggatccacg cctatcgcga atacgactgt 12180 acggatctcg acgtgccgag gctgagccgc gcctttcgga aagtcgtcgc gcggcacgac 12240 atgcttcggg cccacacgct gcccgacatg atgcaggtga tcgagcctaa agtcgacgcc 12300 gacatcgaga tcatcgatct gcgcgggctc gaccggagca cacgggaagc gaggctcgtg 12360 tcgttgcgag atgcgatgtc gcaccgcatc tatgacaccg agcgccctcc gctctatcac 12420 gtcgtcgccg ttcggctgga cgagcggcaa acccgtctcg tgctcagtat cgatctcatt 12480 aacgttgacc taggcagcct gtccatcatc ttcaaggact ggctcagctt ctacgaagat 12540 cccgagacct ctctccctgt cctggagctc tcgtaccgcg attatgtact cgcgctggag 12600 tctcgcaaga agtctgaggc gcatcaacga tcgatggatt actggaagcg gcgcatcgcc 12660 gagctcccac ctccgccgac gcttccgatg aaggccgatc catctaccct gaaggagatc 12720 cgcttccggc acacggagca atggctgccg tcggactcct ggggtcgatt gaagcggcgt 12780 gtcggggagc gcgggctgac cccgacgggc gtcatcctgg ctgcattttc cgaggtgatc 12840 gggcgctgga gcgcgagccc ccggtttacg ctcaacataa cgctcttcaa ccggctcccc 12900 gtccatccgc gcgtgaacga tatcaccggg gacttcacgt cgatggtcct cctggacatc 12960 gacaccactc gcgacaagag cttcgaacag cgcgctaagc gtattcaaga gcagctgtgg 13020 gaagcgatgg atcactgcga cgtaagcggt atcgaggtcc agcgagaggc cgcccgggtc 13080 ctggggatcc aacgaggcgc attgttcccc gtggtgctca cgagcgcgct taaccagcaa 13140 gtcgttggtg tcacctcgtt gcagaggctc ggaactccgg tgtacaccag cacgcagact 13200 cctcagctgc tgctggatca tcagctctac gagcacgatg gggacctcgt cctcgcgtgg 13260 gacatcgtcg acggagtgtt cccgcccgac cttctggacg acatgctcga agcgtacgtc 13320 gtttttctcc ggcggctcac tgaggaacca tggggtgaac aggtgcgctg ttcgcttccg 13380 cctgcccagc tagaagcgcg ggcgagcgca aacgcgacca acgcgctgct gagcgagcat 13440 acgctgcacg gcctgttcgc ggcgcgggtc gagcagctgc ccatgcagct cgccgtggtg 13500 tcggcgcgca agacgctcac gtacgaagag ctttcgcgcc gttcgcggcg acttggcgcg 13560 cggctgcgcg agcaggggcc acgcccgaac acattggtcg cggtggtgat ggagaaaggc 13620 tgggagcagg ttgtcgcggt tctcgcggtg ctcgagtcag gcgcggccta cgtgccgatc 13680 gatgccgacc taccggcgga gcgtatccac tacctcctcg atcatggtga ggtaaagctc 13740 gtgctgacgc agccatggct ggatggcaaa ctgtcatggc cgccggggat ccagcggctg 13800 ctcgtgagcg aggccggcgt cgaaggcgac ggcgaccagc ctccgatgat gcccattcag 13860 acaccttcgg atctcgcgta tgtcatctac acctcgggat ccacagggtt gcccaagggg 13920 gtgatgatcg atcatcgggg tgccgtcaac accatcctgg acatcaacga gcgcttcgaa 13980 atagggcccg gagacagggt gctggcgctc tcctcgctga gcttcgatct ctcggtctat 14040 gatgtgttcg ggatcctggc ggcgggcggt acgatcgtgg tgccggacgc gtccaagctg 14100 cgcgatccgg cgcattgggc agagttgatc gaacgagaga aggtgacggt gtggaactcg 14160 gtgccggcgc tgatgcggat gctcgtcgag cattttgagg gtcgccccga ttcgctcgct 14220 aggtctctgc ggctttcgct gctgagcggc gactggatcc cggtgggcct gcctggcgag 14280 ctccaggcca tcaggcccgg cgtgtcggtg atcagcctgg gcggggccac cgaagcgtcg 14340 atctggtcca tcgggtaccc cgtgaggaac gtcgacctat cgtgggcgag catcccctac 14400 ggccgtccgc tgcgcaacca gacgttccac gtgctcgatg aggcgctcga accgcgcccg 14460 gtctgggttc cggggcaact ctacattggc ggggtcgggc tggcactggg ctactggcgc 14520

Detailed Description Paragraph Table (26):

Arg Gly Arg Gln Pro Lys Arg Ser Gln Gln Gly Gly His Met Glu Lys 20 25 30 Pro Ile Gly Arg Thr Arg Trp Ala Ile Ala Glu Gly Tyr Ile Pro Gly 35 40 45 Arg Ser Asn Gly Pro Glu Pro Gln Met Thr Ser His Glu Thr Ala Cys 50 55 60 Leu Leu Asn Ala Ser Asp Arg Asp Ala Gln Val Ala Ile Thr Val Tyr 65 70 75 80 Phe Ser Asp Arg Asp Pro Ala Gly Pro Tyr Arg Val Thr Val Pro Ala 85 90 95 Arg Thr Arg His Val Arg Phe Asn Asp Leu Thr Glu Pro Glu Pro 100 105 110 Ile Pro Arg Asp Thr Asp Tyr Ala Ser Val Ile Glu Ser Asp Val Pro 115 120 125 Ile Val Val Gln His Thr Arg Leu Asp Ser Arg Gln Ala Glu Asn Ala 130 135 140 Leu Ile Ser Thr Ile Ala Tyr Thr Asp Arg Glu 145 150 155 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 21 <211> LENGTH: 156 <212> TYPE: PRT <213> ORGANISM: Sorangium cellulosum <400> SEQUENCE: 21 Val Arg Arg Ser Arg Trp Gln Met Lys His Val Asp Thr Gly Arg Arg 1 5 10 15 Val Gly Arg Arg Ile Gly Leu Thr Leu Gly Leu Leu Ala Ser Met Ala 20 25 30 Leu Ala Gly Cys Gly Gly Pro Ser Glu Lys Ile Val Gln Gly Thr Arg 35 40 45 Leu Ala Pro Gly Ala Asp Ala His Val Ala Ala Asp Val Asp Pro Asp 50 55 60 Ala Ala Thr Thr Arg Leu Ala Val Asp Val Val His Leu Ser Pro Pro 65 70 75 80 Glu Arg Ile Glu Ala Gly Ser Glu Arg Phe Val Val Trp Gln Arg Pro 85 90 95 Ser Ser Glu Ser Pro Trp Gln Arg Val Gly Val Leu Asp Tyr Asn Ala 100 105 110 Ala Ser Arg Arg Gly Lys Leu Ala Glu Thr Thr Val Pro His Ala Asn 115 120 125 Phe Glu Leu Leu Ile Thr Val Glu Lys Gln Ser Ser Pro Gln Ser Pro 130 135 140 Ser Ser Ala Ala Val Ile Gly Pro Thr Ser Val Gly 145 150 155 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 22 <211> LENGTH: 305 <212> TYPE: PRT <213> ORGANISM: Sorangium cellulosum <400> SEQUENCE: 22 Met Glu Lys Glu Ser Arg Ile Ala Ile Tyr Gly Ala Ile Ala Ala Asn 1 5 10 15 Val Ala Ile Ala Ala Val Lys Phe Ile Ala Ala Ala Val Thr Gly Ser 20 25 30 Ser Ala Met Leu Ser Glu Gly Val His Ser Leu Val Asp Thr Ala Asp 35 40 45 Gly Leu Leu Leu Leu Leu Gly Lys His Arg Ser Ala Arg Pro Pro Asp 50 55 60 Ala Glu His Pro Phe Gly His Gly Lys Glu Leu Tyr Phe Trp Thr Leu 65 70 75 80 Ile Val Ala Ile Met Ile Phe Ala Ala Gly Gly Gly Val Ser Ile Tyr 85 90 95 Glu Gly Ile Leu His Leu Leu His Pro Arg Gln Ile Glu Asp Pro Thr 100 105 110 Trp Asn Tyr Val Val Leu Gly Ala Ala Ala Val Phe Glu Gly Thr Ser 115 120 125 Leu Ile Ile Ser Ile His Glu Phe Lys Lys Lys Asp Gly Gln Gly Tyr 130 135 140 Leu Ala Ala Met Arg Ser Ser Lys Asp Pro Thr Thr Phe Thr Ile Val 145 150 155 160 Leu Glu Asp Ser Ala Ala Leu Ala Gly Leu Thr Ile Ala Phe Leu Gly 165 170 175 Val Trp Leu Gly His Arg Leu Gly Asn Pro Tyr Leu Asp Gly Ala Ala 180 185 190 Ser Ile Gly Ile Gly Leu Val Leu Ala Ala Val Ala Val Phe Leu Ala 195 200 205 Ser Gln Ser Arg Gly Leu Leu Val Gly Glu Ser Ala Asp Arg Glu Leu 210 215 220 Leu Ala Ala Ile Arg Ala Leu Ala Ser Ala Asp Pro Gly Val Ser Ala 225 230 235 240 Val Gly Arg Pro Leu Thr Met His Phe Gly Pro His Glu Val Leu Val 245 250 255 Val Leu Arg Ile Glu Phe Asp Ala Ala Leu Thr Ala Ser Gly Val Ala

260 265 270 Glu Ala Ile Glu Arg Ile Glu Thr Arg Ile Arg Ser Glu Arg Pro Asp 275 280 285 Val Lys His Ile Tyr Val Glu Ala Arg Ser Leu His Gln Arg Ala Arg 290 295 300 Ala 305 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 23 <211> LENGTH: 135 <212> TYPE: PRT <213> ORGANISM: Sorangium cellulosum <400> SEQUENCE: 23 Val Gln Thr Ser Ser Phe Asp Ala Arg Tyr Ala Gly Cys Lys Ser Ser 1 5 10 15 Arg Arg Ile Ala Arg Ser Gly Ser Ala Gly Ala Arg Ala Gly Arg Ala 20 25 30 His Glu Gly Ala Ala Ser Ala Gly Phe Glu Gly Gly Asp Val Met Arg 35 40 45 Lys Ala Arg Ala His Gly Ala Met Leu Gly Gly Arg Asp Asp Gly Trp 50 55 60 Arg Arg Gly Leu Pro Gly Ala Gly Ala Leu Arg Ala Ala Leu Gln Arg 65 70 75 80 Gly Arg Ser Arg Asp Leu Ala Arg Arg Arg Leu Ile Ala Ser Val Ser 85 90 95 Leu Ala Gly Gly Ala Ser Met Ala Val Val Ser Leu Phe Gln Leu Gly 100 105 110 Ile Ile Glu Arg Leu Pro Asp Pro Pro Leu Pro Gly Phe Asp Ser Ala 115 120 125 Lys Val Thr Ser Ser Asp Ile 130 135 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 24 <211> LENGTH: 19 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence universal reverse primer <400> SEQUENCE: 24 ggaaacagct atgaccatg 19 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 25 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence universal forward primer <400> SEQUENCE: 25 gtaaaacgac ggccagt 17 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 26 <211> LENGTH: 28 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence PCR primer NH24 end "B" <400> SEQUENCE: 26 gtgactggcg cctggaatct gcatgagc 28 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 27 <211> LENGTH: 28 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence PCR primer NH2 end "A" <400> SEQUENCE: 27 agcgggagct tgctagacat tctgtttc 28 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 28 <211> LENGTH: 24 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence PCR primer NH2 end "B" <400> SEQUENCE: 28 gacgcgcctc gggcagcgcc ccaa 24 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 29 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence PCR primer pEPO15-NH6 end "B" <400> SEQUENCE: 29 caccgaagcg tcgatctggt ccatc 25 <200> SEQUENCE CHARACTERISTICS: <210> SEQ ID NO 30 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence PCR primer pEPO15H2.7 end "A" <400> SEQUENCE: 30 cggtcagatc gacgacgggc tttcc 25

Other Reference Publication (1):
Bollag, et al., Epothilones, A New Class of Micro-Tubule-stabilizing Agnets with a Taxol-like Mechanism of Action, Cancer Research, 55, 2325-2333, Jun. 1995.

Other Reference Publication (3):
Nicolaou, et al., Chemical Biology of Epothilones, Angew. Chem. Int. Ed., 1998, 37, 2014-2045.

Other Reference Publication (4):
Schupp, et al., Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from Amycolatopsis mediterranei, FEMS Microbiology Letters, 159, 1998, 201-207.

Other Reference Publication (11):
Molnar et al., Gene 169(1):1-7 (1996).

Other Reference Publication (12):
Aparicio et al., Gene 169(1)9-16 (1996).

CLAIMS:

1. An isolated nucleic acid fragment comprising a nucleotide sequence that encodes at least one polypeptide required for the biosynthesis of epothilone, wherein the complement of said nucleotide sequence hybridizes to a sequence selected from the group consisting of: nucleotides 43524-54920 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, nucleotides 51534-52657 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, and nucleotides 54540-54758 of SEQ ID NO:1, under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5% SDS for 20 minutes at 65.degree. C.

2. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 1.

3. A recombinant vector comprising a chimeric gene according to claim 2.

4. A recombinant host cell comprising a chimeric gene according to claim 2.

8. An isolated nucleic acid fragment according to claim 1, wherein said polypeptide comprises a .beta.-ketoacyl-synthase domain and wherein the complement of said nucleotide sequence hybridizes to a sequence selected from the group consisting of: nucleoticles 43626-44885 of SEQ ID NO:1 and nucleotides 48087-49361 of SEQ ID NO:1, under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5%

'SDS for 20 minutes at 65.degree. C.

9. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 8.

10. A recombinant vector comprising a chimeric gene according to claim 9.

11. A recombinant host cell comprising a chimeric gene according to claim 9.

15. An isolated nucleic acid fragment according to claim 1, wherein said polypeptide comprises an acyltransferase domain and wherein the complement of said nucleotide sequence hybridizes to a sequence selected from the group consisting of: nucleotides 45204-46166 of SEQ ID NO:1 and nucleotides 49680-50642 of SEQ ID NO:1, under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5% SDS for 20 minutes at 65.degree. C.

16. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 15.

17. A recombinant vector comprising a chimeric gene according to claim 16.

18. A recombinant host cell comprising a chimeric gene according to claim 16.

22. An isolated nucleic acid fragment according to claim 1, wherein said polypeptide comprises a dehydratase domain and wherein the complement of said nucleotide sequence hybridizes to nucleotides 50670-51176 of SEQ ID NO:1 under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5% SDS for 20 minutes at 65.degree. C.

23. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 22.

24. A recombinant vector comprising a chimeric gene according to claim 23.

25. A recombinant host cell comprising a chimeric gene according to claim 23.

29. An isolated nucleic acid fragment according to claim 1, wherein said polypeptide comprises a methyltransferase domain and wherein the complement of said nucleotide sequence hybridizes to nucleotides 51534-52657 of SEQ ID NO:1 under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5% SDS for 20 minutes at 65.degree. C.

30. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 29.

31. A recombinant vector comprising a chimeric gene according to claim 30.

32. A recombinant host cell comprising a chimeric gene according to claim 30.

36. An isolated nucleic acid fragment according to claim 1, wherein said polypeptide comprises a .beta.-ketoreductase domain and wherein the complement of said nucleotide sequence hybridizes to a sequence selected from the group consisting of: nucleotides 46950-47702 of SEQ ID NO:1 and nucleotides 53697-54431 of SEQ ID NO:1, under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5% SDS for 20 minutes at 65.degree. C.

37. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 36.

38. A recombinant vector comprising a chimeric gene according to claim 37.

39. A recombinant host cell comprising a chimeric gene according to claim 37.

43. An isolated nucleic acid fragment according to claim 1, wherein said polypeptide comprises an acyl carrier protein domain and wherein the complement of said nucleotide sequence hybridizes to a sequence selected from the group consisting of: nucleotides 47811-48032 of SEQ ID NO:1 and nucleotides 54540-54758 of SEQ ID NO:1, under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5% SDS for 20 minutes at 65.degree. C.

44. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 43.

45. A recombinant vector comprising a chimeric gene according to claim 44.

46. A recombinant host cell comprising a chimeric gene according to claim 44.

50. An isolated nucleic acid fragment comprising a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence selected from the group consisting of: SEQ ID NO:6, amino acids 35-454 of SEQ ID NO:6, amino acids 561-881 of SEQ ID NO:6, amino acids 1143-1393 of SEQ ID NO:6, amino acids 1430-1503 of SEQ ID NO:6, amino acids 1522-1946 of SEQ ID NO:6, amino acids 2053-2373 of SEQ ID NO:6, amino acids 2383-2551 of SEQ ID NO:6, amino acids 2671-3045 of SEQ ID NO:6, amino acids 3392-3636 of SEQ ID NO:6, and amino acids 3673-3745 of SEQ ID NO:6.

51. An isolated nucleic acid fragment according to claim 50, wherein said nucleotide sequence is selected from the group consisting of: nucleotides 43524-54920 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1, nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, nucleotides 51534-52657 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, and nucleotides 54540-54758 of SEQ ID NO:1.

52. An isolated nucleic acid fragment according to claim 50, wherein said polypeptide comprises a .beta.-ketoacyl-synthase domain comprising an amino acid sequence selected from the group consisting of: amino acids 35-454 of SEQ ID NO:6 and amino acids 1522-1946 of SEQ ID NO:6.

53. An isolated nucleic acid fragment according to claim 52, wherein said nucleotide sequence is selected from the group consisting of: nucleotides 43626-44885 of SEQ ID NO:1 and nucleotides 48087-49361 of SEQ ID NO:1.

54. An isolated nucleic acid fragment according to claim 50, wherein said polypeptide comprises an acyltransferase domain comprising an amino acid sequence selected from the group consisting of: amino acids 561-881 of SEQ ID NO:6 and amino acids 2053-2373 of SEQ ID NO:6.

55. An isolated nucleic acid fragment according to claim 54, wherein said nucleotide sequence is selected from the group consisting of: nucleotides 45204-46166 of SEQ ID NO:1 and nucleotides 49680-50642 of SEQ ID NO:1.

56. An isolated nucleic acid fragment according to claim 50, wherein said polypeptide comprises a dehydratase domain comprising amino acids 2383-2551 of SEQ ID NO:6.

57. An isolated nucleic acid fragment according to claim 56, wherein said nucleotide sequence is nucleotides 50670-51176 of SEQ ID NO:1.

58. An isolated nucleic acid fragment according to claim 56, wherein said polypeptide comprises a methyltransferase domain comprising amino acids 2671-3045 of SEQ ID NO:6.

59. An isolated nucleic acid fragment according to claim 58, wherein said nucleotide sequence is nucleotides 51534-52657 of SEQ ID NO:1.

60. An isolated nucleic acid fragment according to claim 50, wherein said polypeptide comprises a .beta.-ketoreductase domain comprising an amino acid sequence selected from the group consisting of: amino acids 1143-1393 of SEQ ID NO:6 and amino acids 3392-3636 of SEQ ID NO:6.

61. An isolated nucleic acid fragment according to claim 60, wherein said nucleotide sequence is selected from the group consisting of: nucleotides 46950-47702 of SEQ ID NO:1 and nucleotides 53697-54431 of SEQ ID NO:1.

62. An isolated nucleic acid fragment according to claim 50, wherein said polypeptide comprises an acyl carrier protein domain comprising an amino acid sequence selected from the group consisting of: amino acids 1430-1503 of SEQ ID NO:6 and amino acids 3673-3745 of SEQ ID NO:6.

63. An isolated nucleic acid fragment according to claim 62, wherein said nucleotide sequence is selected from the group consisting of: nucleotides 47811-48032 of SEQ ID NO:1 and nucleotides 54540-54758 of SEQ ID NO:1.

64. A chimeric gene comprising a heterologous promoter sequence operatively linked to a nucleic acid fragment according to claim 50.

65. A recombinant vector comprising a chimeric gene according to claim 64.

66. A recombinant host cell comprising a chimeric gene according to claim 64.

70. An isolated polypeptide required for the biosynthesis of epothilone, wherein said polypeptide comprises an amino acid sequence encoded by a nucleotide sequence whose complement hybridizes to a sequence selected from the group consisting of: nucleotides 43524-54920 of SEQ ID NO:1, nucleotides 43626-44885 of SEQ ID NO:1, nucleotides 45204-46166 of SEQ ID NO:1,

nucleotides 46950-47702 of SEQ ID NO:1, nucleotides 47811-48032 of SEQ ID NO:1, nucleotides 48087-49361 of SEQ ID NO:1, nucleotides 49680-50642 of SEQ ID NO:1, nucleotides 50670-51176 of SEQ ID NO:1, nucleotides 51534-52657 of SEQ ID NO:1, nucleotides 53697-54431 of SEQ ID NO:1, and nucleotides 54540-54758 of SEQ ID NO:1, under conditions of hybridization at 65.degree. C. for 36 hours and washing 3 times at high stringency with 0.1.times.SSC and 0.5% SDS for 20 minutes at 65.degree. C.